THE UNIVERSITY OF ALBERTA

CONCEPT IDENTIFICATION AS A FUNCTION OF PROPORTION OF POSITIVE INSTANCES AND CONCEPT SIZE

by

DONALD GEORGE WARGO

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

EDMONTON, ALBERTA

AUGUST, 1966

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Concept Identification as a Function of Proportion of Positive Instances and Concept Size" submitted by Donald George Wargo in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

The present study investigated the effects of four proportions of positive instances (1/8, 2/8, 3/8, and 4/8) and three levels of concept size (one, three, and five nonredundant relevant binary stimulus dimensions) in a conjunctive concept identification task. Visual stimuli were presented to 192 university undergraduate Ss, each serving once in a problem consisting of 96 instances. A two-choice response system was used with 100% informative feedback. The primary measure of performance employed was errors per block of 16 instances.

The results showed that mean total errors increased as the proportion of positive instances increased in the range 1/8 to 3/8, but that the 4/8 condition led to fewer errors than any of the other three. It was also found that errors increased as a negatively accelerated function of size. Proportion of positive instances did not interact with size.

The results, interpreted in relation to other research, suggest that Ss utilized differential category frequencies as a basis for responding to reduce errors when this was possible. The results have serious implications for concept identification research, and for an understanding of the process of concept learning.

Acknowledgments

The writer wishes to express special gratitude to Dr. Willard N. Runquist for his advice and guidance in the planning, execution, and writing of this thesis. The advice of Dr. W. A. S. Smith, Dr. William A. Blanchard, Dr. Donald Spearman, and Dr. Stanley J. Rule, offered at various stages, is also greatly appreciated.

The writer is indebted to his wife, Lynne D. Wargo, for doing all the typing, and for her patience and encouragement throughout.

Donald George Wargo

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The primary purpose of the present research investigation is to determine the effect on categorization performance of different proportions of positive and negative instances of a concept when the difference in category frequencies permits probabilistic responding. The importance of this investigation is twofold. First, if different proportions of positive and negative instances lead to differential performance in such a situation, then this constitutes a finding of some methodological importance for the researcher and the educator. For the researcher it suggests that results may vary as a function of the constant frequencies of positive and negative instances employed. For the educator, more efficient communication of concepts may result from the findings. Second, if probabilistic responding produced by this differential may be inferred, then the study may yield information about the process of concept identification, and may have implications for an understanding of other related cognitive activities when probabilistic responding is possible in the situation. This, too, may yield important information for the researcher, enabling him to identify, and perhaps control, yet another variable which may affect the isolation of learning per se as opposed to performance.

In the typical concept identification experiment _S_s are presented with a number of stimuli which they must categorize. These stimuli are either exemplars of what the concept _is_, i.e., positive instances, or of what the concept _is not_, i.e., negative instances. It is through these two types of instances and informative feedback regarding the correct classification of each that _S_ attempts to

inductively arrive at what the concept is so that he may either specify it or proceed to correctly categorize remaining instances. Because the concept is conveyed to S via these two types of instances, the relative efficiency of each type is important to an understanding of conceptual behavior.

Previous studies on types of instances have employed various methodologies, and for the most part have investigated only a small number of the possible combinations of positive and negative instances. As such, knowledge of their differential effects on performance is not complete. In view of the growing emphasis on and proliferation of research on conceptual behavior, the effects of positive and negative instances should be thoroughly investigated under the various methodologies typically employed. The present study investigates the effects on performance of proportions of positive vs. negative instances in a task involving a fixed number of stimulus presentations.

The manipulation of proportions of these two types of instances is accomplished by the use of differential stimulus frequencies in a two-choice categorization task. It is suggested that one way in which the differential proportions may be effective in enhancing performance is by facilitating probability learning. Learning of the differential frequencies may enable Ss to improve their performance in a concept identification situation over and above any concept learning that is taking place, and this performance enhancement may vary as a function of the differential frequencies of the stimulus classes.

In addition to proportions of positive and negative instances, the present study investigates the effect of concept size, or number

of nonredundant stimulus dimensions, on concept identification performance. This variable has yielded conflicting results in past research, and the present study seeks to clarify its effect in a fixed trials situation in conjunction with the variable of proportion of positive (or negative) instances.

The interaction between these two major variables is also of interest since it would seem likely that if, as suggested, Ss utilize the differential frequencies as a basis for responding, then they should do so to a greater extent as concept size is increased and the task becomes more difficult. Since the potential benefit of probabilistic responding varies with the extent of the frequency differential, then the combination of these two variables might well result in an interaction effect.

Concepts and Conceptual Behavior

In recent years there has been an increase in the amount of interest shown in the learning, identification, and formation of concepts. This emphasis results largely from the fact that the study of conceptual behavior is one possible approach to the investigation of cognitive functioning, and more specifically to the investigation of thinking. Underwood (1952) states that the major approaches to the investigation of thinking are found in problem solving, reasoning, creativity, and concept formation research. In fact, one of the landmarks in the area of concept learning by Bruner, Goodnow, and Austin (1956) is entitled _A Study of Thinking_. More recently Hunt (1962) states "We feel intuitively that concepts are things used in thoughts",

while Bower and Trabasso (1964) point out that the basic units of man's language and thoughts are concepts.

The term "concept" has been given various definitions by writers in the field. For example, Bruner, Goodnow, and Austin define a concept as "A network of sign-significate inferences by which one goes beyond a set of observed criterial properties exhibited by an object or event to the class identity of the object or event in question, and thence to additional inferences about other unobserved properties of the object or event". Heidbreder (1946) defines a concept as "A logical construct which, through signs or symbols or both, is transferable from situation to situation and communicable from person to person".

These definitions are somewhat theoretical and abstract. The definition by Bruner et al. includes not only a very general definition of a concept, but actually emphasizes the use of concepts that have been acquired. Heidbreder's, too, includes only a very general definition of the term, but additionally emphasizes the condition that it be communicable and transferable to other situations. But neither of these concretely specifies what a concept is in behavioral terms, and hence they do not lend themselves well to operational utilization. Kendler (1961), on the other hand, provides a simpler, more concrete definition which lends itself to operational usage, and essentially defines the behavioral phenomenon to be studied. She defines a concept as "A common response to dissimilar stimuli". In its parsimony, however, this definition is perhaps too restrictive. It excludes, for example, categorization tasks involving only two

categories where no irrelevant dimensions are present, but where multiple relevant dimensions are present. This situation certainly cannot be considered a simple discrimination task (except insofar as all concept learning tasks might be regarded as such), nor can it be regarded as a paired associates task, except in that one of the types of stimuli constitutes a class by itself.

Further differences and similarities between these and other definitions of the term would be interesting to consider in greater detail. Hunt (1962) considers some of these definitions at length. Such differences and similarities are irrelevant, however, for the purposes of the present paper, and any definition which includes or implies that a concept is an ideational class will suffice.

There is no disagreement on the fact that conceptual behavior involves classification or categorization. Indeed, Bruner et al. suggest that all cognitive activity is dependent upon or involves the categorization process. They state that categorizing accomplishes several things for the human organism. Briefly, these might be stated as follows: (1) The organism reduces the complexity of its environment by categorizing discriminably different events as being equivalent; (2) Categorizing constitutes the means by which objects in the world are identified; (3) The necessity of constant learning is reduced by the establishment of categories based upon sets of defining attributes; (4) Categories provide a direction for instrumental activity; and (5) Categorizing enables us to order and relate classes of events.

The <u>S</u>'s task in the typical concept identification experiment

is to classify different stimuli, and to discover through trial and error the rules whereby objectively different stimuli are assigned to the same category. This class of behavior usually does not involve the acquisition of novel attributes since the perceptual discriminations and names of attributes relevant to the solution of the problem are already a part of the person's behavioral repertoire.

Terminology

In order to facilitate communication, explanation or definition of some frequently used terms in the area of concept learning research, as well as in the present paper, are presented.

Experimental studies of concept learning involve different forms of categorizing behavior. If an array of stimuli is presented to *S* and he is instructed to divide the stimuli into categories of his own choosing, then it is most appropriate to refer to this process as concept *formation*. Perhaps more frequently, however, serial presentations of stimuli are used and *S* is instructed to categorize on the basis of some concept that *E* has in mind which *S* is to discover. In this latter case it is more appropriate to use the term concept *identification*. In this process, when *S* correctly identifies the rules for categorization it is said that *S* *attains* the concept. Attainment is usually operationally defined.

The term *dimension* refers to some aspect of the stimuli which may vary between two or more *values*, or *levels*, on that dimension. *Binary dimensions* are frequently used, and these appear at two different levels in a series of stimuli. A dimension appears at only

one of its levels in a given stimulus. Thus, a given stimulus pattern will involve a given value of each of one or more dimensions. For example, if color is a binary dimension, it would appear at either of two different values or levels, such as red and green, in different stimulus presentations. The dimensions may be subdivided into _relevant dimensions_ and _irrelevant dimensions_. A relevant dimension is one which is essential to the concept in question. For example, in the concept "two squares", number is relevant, as is shape. An irrelevant dimension is one which appears at its different levels in a series of examples of the concept, but which is nonessential to the concept in question. For example, in the concept "two squares", color would be irrelevant if it varied between red and green in different examples of the concept. A single stimulus pattern is called an _instance_. Instances may be either _positive_ or _negative_. A positive instance of the concept is any stimulus which contains the certain specified (positive) levels of all relevant dimensions. A negative instance of the concept is any stimulus which does not contain all positive levels of all relevant dimensions.

Research Perspective

While human conceptual behavior has long been of interest to psychologists, it was not until the early 1950's that it became a major area of laboratory experimentation. Since then the number of experimental investigations has increased considerably. Reviews of the research literature on conceptual behavior by Tracy Kendler (1961), and more recently by Bourne (1966), clearly indicate the broad scope

encompassed by these investigations.

H. Kendler (1964) observes that this research has taken two major directions. One of these consists of efforts to discover systematic relationships between stimulus events and a common response. In this group, the emphasis is placed primarily upon the stimulus-response relationship. The other major direction is focused on the mediational mechanism responsible for conceptual behavior. In this case, the internal cue instead of the association is the main focus.

It is with the former of these two directions that the present study is associated, in that it investigates the effect on performance of certain stimulus variables using methodological procedures designed for research of this type. Among the variables investigated within this overall approach have been stimulus redundancy, feedback, stimulus complexity, and amounts of information, to mention just a few.

Before proceeding to consider research related to the specific variables of concern in the present paper, two early research investigations will be briefly described which constitute prototypes for much of the work done since.

Laboratory research in concept learning dates back to Hull's (1920) classic study using Chinese characters as stimuli. Hull used six packs of 12 cards each with Chinese characters drawn on them. Subjects were shown each card in a pack for five seconds. Simultaneously E made a nonsense sound such as "oo", or "yer". Each symbol that was paired with a particular sound had some component of the symbol in common with other symbols paired with that same sound. After E went through the first pack of 12 cards making sounds with

each symbol, the S's task was to see if he too could make the correct verbal response as each card was shown in succeeding packs. Hull then measured the number of trials to learn the concept. Most importantly, he measured the percentage of correct responses on the first trial of each pack of cards. The results on this latter measure showed a negatively accelerated function ranging from 27% on pack number two, to 56% on the last pack. An interesting aspect of the results was that some Ss reported that they did not see the component of the symbol designated as the concept, but they were able to determine that the symbols going with the same sound looked somewhat alike, or that they had a feeling for what the concept was without any specific awareness of the essential component.

In the 1940's, Heidbreder (1946, 1947) carried out a series of experiments similar to Hull's. She showed her Ss 15 series of drawings on a memory drum, each series consisting of nine pictures. Each picture was paired with a nonsense name, and the S's task was to anticipate the nonsense name whenever possible. Each series was learned to a criterion of two perfect trials. Her stimuli contained pictures of concrete objects, spatial forms, and abstract numbers. For example, a tree was always a "mulp", a circle form was always a "fard", and a stimulus configuration containing six figures was always a "mank". On the basis of her results she concluded that concepts involving concrete objects are the most easily learned, the next easiest are spatial forms, and the most difficult are concepts involving numbers.

Focus of the Present Investigation

Laboratory research on conceptual behavior often begins with the observation that some variable appears to be of importance in such functioning. One example of such a research program has to do with the effects of the number of variable dimensions associated with a class of objects. Considering for a moment only those dimensions which constitute an integral part of a concept, i.e., relevant dimensions, it would seem likely that the greater the number of these necessary to define a concept the harder it will be for a person to identify or learn that concept. When only one such relevant dimension is involved, the task is a simple discrimination task. But when more nonredundant relevant dimensions are necessary to define the concept, then the task becomes more difficult. The question arises as to how much harder it becomes with the addition of more and more relevant dimensions. The answer to this question is not entirely clear in that research on amounts of relevant information, or number of relevant dimensions, have yielded partially conflicting results. The present study attempts partly to focus on this issue, referring to this variable as "concept size". Certainly any general understanding of how people learn concepts will have to consider any effects of concept size since it is known that the variable makes a difference in concept learning, and it is obvious that concepts outside and inside the laboratory vary in terms of their numbers of relevant dimensions.

Another such variable of importance to a general understanding of concept learning has to do with the way in which concepts are

transmitted to the person. It is possible to convey a concept in terms of exemplars of what the concept _is_ _not_ as well as by exemplars of what it _is_. The results of previous studies which focused on types of instances would lead to the general conclusion that positive instances are more effective in transmitting a concept. A previous study (Wargo, 1960) which varied the number of relevant stimulus dimensions, but which held constant a particular proportional division of positive and negative instances, produced results which appeared to be partly a function of the constant level of the proportions of positive and negative instances employed. From this it appeared that _S_s may respond on the basis of learned probabilities of the correctness of responses and use this information to enhance their performance. It is important, therefore, to determine what effect the opportunity for probabilistic responding might have on performance since in categorization tasks this opportunity might often be present due to the nature of the task involved. The experiment reported here addresses itself to this by varying the proportion of each of these two types of exemplars, or instances, in an effort to see what effect this has on performance. If it is demonstrated that the opportunity for probabilistic responding does have an effect, then this will have important implications for methodology in concept identification research.

Relevant Research

A. _Amounts of Information_: One way of classifying concept learning tasks is in terms of the number of stimulus dimensions that

vary. A convenient method of quantifying concept problems is in terms of the information theory concept of "bits" of information (Miller, 1953). When bi-level dimensions are used, as is the case in the present study, the number of bits of information contained in the problem is equal to the number of these binary dimensions. The total amount of information may be further subdivided into the amount of relevant information and the amount of irrelevant information contained in the problem. The amount of irrelevant information refers to the degree of "complexity" of the problem. The amount of relevant information, or the number of relevant binary dimensions involved, is sometimes known as "concept size".

Systematic investigation of amounts of information began with the paper by Archer, Bourne, and Brown (1955). Their problems contained two bits of relevant information, and up to five bits of irrelevant information. They ran Ss to a criterion and found that task difficulty increased as the amount of irrelevant information was increased, as indicated by number of errors and number of trials to the attainment of the concept. Performance deteriorated as a positive exponential function of the amount of irrelevant information.

This work led to a series of investigations of visual concept learning in which task complexity was involved. The studies by Brown and Archer (1956), Bourne (1957), Bourne and Pendleton (1958), Walker and Bourne (1961), and Wargo (1960) all found similar results to those of Archer et al. in that performance was found to deteriorate as a function of the amount of irrelevant information contained in the problem. Some studies in this series contained fixed numbers of trials

and in others Ss were run to a criterion. However, unlike Archer et al. all found the relationship between performance and task complexity not to deviate from linearity. One additional investigation by Bulgarella and Archer (1962) employed auditory stimuli entirely. Subjects were run to a criterion and their results were consistent with the findings of the majority of visual studies in that only the linear component of the trend was significant.

Some of these investigations also varied amount of relevant information. The experiment by Walker and Bourne involved one, two, and three bits of both relevant and irrelevant information. Subjects were run to a criterion. The number of response categories was varied depending upon the amount of relevant information contained in the problem. The number of categories into which stimuli were placed was 2^n, where n was the number of relevant dimensions. Therefore, there were two, four, and eight categories, respectively. In addition to finding that performance deteriorated as a linear function of amount of irrelevant information, they report that errors to criterion increased as a positively accelerated function of the amount of relevant information.

Bulgarella and Archer also varied amounts of both relevant and irrelevant information. One, two, and three bits of relevant information, along with three levels of irrelevant information, were employed in a two category response situation. Their results showed that on all measures--number of errors, trials, and time scores--performance worsened as the amount of relevant information was increased. The effect of amount of relevant information was found to

be linear, unlike the positively accelerated function found by Walker and Bourne.

Wargo's study used zero, one, two, three, and four bits of irrelevant information, and one, two, three, and four bits of relevant information with visual concept stimuli. As in the Bulgarella and Archer study, the stimuli were placed in one of two categories. However, Wargo's procedure differed in that he ran Ss for a fixed number of trials. In addition to finding that performance deteriorated as a linear function of amount of irrelevant information, it was found that as the amount of relevant information was increased, performance deteriorated as a negatively accelerated function.

The difference in findings between the studies by Walker and Bourne, Bulgarella and Archer, and Wargo regarding relevant information should be emphasized. To summarize, all three found performance to deteriorate as a function of amount of relevant information. However, Walker and Bourne found this function to be positively accelerated, Bulgarella and Archer found it to be linear, and Wargo found it to be negatively accelerated. Since Walker and Bourne used a considerably more complicated response-categorization system than the other two studies, it is likely that the positively accelerated increase in performance decrement might be explained by the fact that they increased response uncertainty as they increased amount of relevant information. The investigations carried out by Bulgarella and Archer and by Wargo were similar in that they both involved the same differential category frequencies. They did differ, however, in terms of the sensory modality to which stimuli were presented and in the dependent variable

measures employed. It is conceivable that the difference in the form of the performance functions could result from the differences in sensory modality. Further investigation of this variable would be necessary.

Wargo suggests that a more likely explanation is to be found in the differences between measures and consequent task focus inherent in a fixed trials situation as compared with a trials to criterion task. Since both of these studies employed 25% positive and 75% negative instances, the opportunity existed for probabilistic responding based on this frequency differential. Subjects may well have learned the respective probabilities associated with the two response categories in either study, but they would have been more apt to utilize this information to improve their performance in the case of the fixed trials situation, where the primary focus of the task was upon making correct responses. In the trials to criterion situation, Ss were possibly focused upon problem solution to the relative neglect of errors made in the process. In addition, if probabilistic responding were utilized, this would have been reflected directly in the performance measure employed in the fixed trials situation, but would not have been as clearly reflected in either of the two primary measures employed in the trials to criterion procedure. The effect which probabilistic responding would be expected to have upon performance would be to reduce the proportion of errors, or error rate. This is essentially the measure that was employed in the Wargo study, namely, errors per 96 trials. The errors to criterion measure obtained in the Bulgarella and Archer study does not

directly reflect error rate, since number of trials involved was not held constant, and this measure would be expected to covary with the trials to criterion measure since the two were not independent. Since the trials to criterion measure does not reflect error rate, it should be unaffected by probabilistic responding. Since it was also suggested that as task difficulty increased, the incidence and extent of probabilistic responding would also increase, this might well produce a negatively accelerated error function where such responding was being reflected.

Wargo suggests that his Ss may have reduced their error rates by pressing the lever representing negative instances all the time, i.e. probability maximizing, while still gaining the same information leading to the correct solution of the problem as they would have by varying their responses. This explanation was further supported by the fact that the level of errors made by Ss as amount of relevant information increased appeared to be reaching a ceiling at approximately 25% errors.

B. Positive and Negative Instances: A series of studies was carried out by Smoke (1932, 1933) on the relative effectiveness of positive and negative instances in concept learning. This series compared concept learning using arrays of all positive instances as well as mixed arrays of equal numbers of positive and negative instances. In addition, simultaneous and serial modes of presentation were investigated. Using the amount of time taken to learn the concept, Smoke found no statistically significant differences between all positive instances and mixed arrays. That is, he found no evidence

of faster learning by one method or the other. He did find, however, that in simultaneous presentation, Ss tended to prefer mixed arrays to all positive arrays. He also found in the case of serial presentation that negative instances tended to confuse some Ss. Whereas, in simultaneous presentation, negative instances permitted contrast with the positive instances and therefore seemed to make learning somewhat easier, even though it did not show up in the form of faster learning. In brief, Smoke pointed out that while negative instances did not make for speed of learning, they tended to facilitate accuracy, especially in the case of more difficult concepts. He suggested that negative instances may assist concept learning in that they may prevent the learner from coming to erroneous conclusions while still in the midst of the learning process.

Smoke's experiments led to a series of studies by Hovland and Weiss (1953). They investigated differences in difficulty and assimilation of material presented in the form of positive and negative instances. Since the same concept may be transmitted by either a given number of positive instances or a given number of negative instances, Hovland and Weiss employed the minimum number of instances necessary in order to adequately specify the concept, such that any difference in learning the concept could then be attributed to differences in the difficulty of assimilating information contained in positive instances as compared with information contained in negative instances. Their procedures involved both successive presentation and simultaneous presentation in order to compare these, as well. In addition, Ss were told beforehand the total number of dimensions which would be involved

in each problem, the number of relevant and irrelevant dimensions involved, the number of values for each dimension, and the number of correct values on each relevant dimension. They were also given practice problems with error correction. Then they were presented with the minimum number of instances necessary to completely specify the concept. Results showed that the correct concept was attained by a higher percentage of Ss when the concept was transmitted by all positive instances than by all negative instances, with mixed positive and negative instance presentations falling between. It was also shown that more Ss were able to attain concepts presented in the form of all negative instances when the instances were displayed simultaneously than when they were presented successively. Hovland and Weiss pointed out, however, that under appropriate conditions, over half the Ss were able to attain the correct concept exclusively on the basis of negative instances in the minimum number of instances necessary to specify the concept.

A very recent experiment by Schvaneveldt (1965) varied the probability of positive instances along with the amount of relevant information contained in the problem. He used three levels of probability of positive instances, .125, .250, and .500. Number of relevant stimulus dimensions were one, two, three, and four. These two variables were combined in a factorial design and Ss were run to a criterion of 32 consecutive correct responses in a dichotomous classification situation. Schvaneveldt found that the overall effect of probability of positive instances was significant. Successive halving of the probability of positive instances led to a linear increment in

trials to criterion. Generally, as the probability of positive instances increased, the mean number of trials to attain criterion decreased. The interaction between number of relevant dimensions and probability of positive instances was not significant.

Freibergs and Tulving (1961) investigated changes in performance across a series of visual concept problems presented entirely in the form of either all positive or all negative instances. All concepts could be conveyed by four instances of either type. Subjects were assigned to either the positive or negative instance condition and presented with twenty successive problems. Their task was to verbalize the concept. Mean time to solution was used as the performance measure, with a limit of 3.5 minutes. Their results showed clearly that time to solution in the positive instance condition was very superior to performance in the negative instance condition on early problems. The difference gradually decreased, however, until there was practically no difference by the end of the series of twenty problems. This study supports the assertion that learning is superior when the concept is presented in the form of all positive instances. It adds the qualification, however, that this applies primarily to naive Ss. These results suggest that the usual positive vs. negative instance findings may be attributable to pre-experimental experience in concept learning. It further suggests that training with negative instances might reduce or eliminate differences between these two types of instances.

One additional recent report which should be cited here is a paper by Mandler, Cowan, and Gold (1964). It is relevant to the study

of types of instances since these researchers investigated the matching of response frequencies to the frequencies of the two types of instances. More specifically, they investigated concept learning performance in a situation involving two-thirds positive and one-third negative instances. Level of difficulty of the concepts was also varied. They sought to demonstrate that Ss will learn the frequencies of responses and will actually probability match prior to concept attainment. Their results supported this hypothesis, and demonstrated that responding based upon probability learning varied inversely with task difficulty. That is, they found that the more difficult the concept problem, the lower was the level of probability responding. They suggested that the kinds of hypotheses that Ss developed based on concept information diverged from the probability structure of the task and thus tended to lower matching behavior. In other words, as Ss developed hypotheses concerning the concept itself, their responding deviated from hypotheses based upon the frequency of input, or probability matching hypotheses. This then led to deviation from asymptotic levels based upon frequencies alone.

Mandler et al. concluded that concept learning and probability matching go on concurrently in such problems. They pointed out that probability matching results in some improvement in event matching performance i.e., responses based on concept learning, and either probability matching or event matching will yield better than chance success. When no other alternative is available to Ss, probability matching occurs, leading to improved performance. When both concept learning and probability matching are available, they will match the

probability structure as well as engaging in more efficient event matching behavior.

<u>The Present Investigation</u>

The present research was stimulated by the inconsistency among some of the previously mentioned studies with respect to the effect of number of relevant stimulus dimensions on concept identification performance. Walker and Bourne (1961), Bulgarella and Archer (1962), and Wargo (1960) each found different performance functions as size was increased. Walker and Bourne's positively accelerated function very likely resulted from their confounding of response uncertainty with size. The difference between Wargo's results and Bulgarella and Archer's results may well be attributable to a difference in methodology. This interpretation suggests that, in a task involving a fixed number of trials and unequal category frequencies, <u>S</u>s may use the frequency differential to reduce errors, and may do so to a greater extent as concept size increases. This suggestion is consistent with the findings of Mandler et al., reported since the inception of the present study, except insofar as they concluded that difficulty and probability responding were inversely related. The scope of their investigation was limited, however, in that they used only one frequency distribution of positive and negative instances.

It would seem likely that <u>S</u>s would be able to utilize this frequency differential with greater benefit as the differential is increased. Thus, if 7/8 of the instances were positive and 1/8 negative, then <u>S</u>s should be able to minimize errors to a greater extent

on the basis of probability learning alone than if this differential involved 5/8 positive instances and 3/8 negative. Therefore, the greater the differential the greater would be the benefit to pre-attainment performance if the probability learning were maximally utilized. Also, the greater the differential, the greater should be the obviousness of that differential.

In this example, however, if it were shown that Ss given problems containing 7/8 positive and 1/8 negative instances were superior to Ss given problems with 5/8 positive and 3/8 negative, the interpretation of these results would be open to debate. Performance superiority on the problem with the greatest frequency differential would support the probability learning suggestion, but it would also be consistent with what is known about the effects of proportions of positive and negative instances. That is, on the basis of an interpolation and generalization of the results of the studies by Hovland and Weiss (1953) and the results reported by Schvaneveldt (1965), it is likely that performance on concept learning tasks is enhanced by greater numbers of positive instances. As a result, an alternative explanation would be available, and the presence of any pre-attainment probability learning would be in question. This potential problem is avoided by reversing the proportions, or by studying the opposite end of the continuum of frequency differentials, such that there are fewer positive instances (or else an equal number) compared with negative instances. In this arrangement the condition which lends itself to the greatest extent to probability learning will be at one extreme of the differential frequency conditions, and the condition which should

facilitate maximum concept learning will be at the other extreme. The actual differential conditions employed in the present study are as follows: 1/8 positive and 7/8 negative instances; 2/8 positive and 6/8 negative; 3/8 positive and 5/8 negative; and 4/8 positive and 4/8 negative.

While probability learning studies have demonstrated that <u>S</u>s learn to respond on the basis of past frequencies, they typically involve situations where there is no solvable problem. The present study attempts to determine the effects on performance when differential category frequencies may be used as a basis for pre-attainment responding in a solvable concept identification task.

Specifically, the present research was designed to investigate the effect of different proportions of positive and negative instances in the range of differential frequencies indicated above, for a fixed number of instances, over a range of problems of different sizes. The results should lend support to either of two predictions which might be made regarding this variable. If performance improves as proportion of positive instances increases, then the finding will be consistent with other research on positive and negative instances. The superiority of positive instances will then be shown to apply under the differences in methodology employed in the present study i.e., fixed number of instances, errors as measure, etc. The likelihood of probabilistic responding in such a situation would then be contraindicated. On the other hand, if performance is better with fewer positive instances and a greater frequency differential, then this should constitute evidence for probabilistic responding as an

aid to performance in such a situation, and as such, will constitute a demonstration of a phenomenon which may be important to both research methodology and an understanding of the concept identification process.

If learned frequencies are used as suggested, it is possible that they will be used to a greater extent as the concept becomes more difficult to attain since Ss will be instructed to make the maximum number of correct responses. This is the usual instruction in concept identification studies where Ss are run for a fixed number of instances. As such, the variable of concept size will also be included to determine whether proportion interacts with size. Size is of interest in its own right, however, in light of the inconsistency in the results of studies on this variable. Its inclusion will thus constitute a partial replication of the earlier study by Wargo (1960).

Changes in performance over the course of the problem will also be studied to facilitate investigation of these major variables, and because relatively little is known about performance changes over the course of concept problems (Hunt, 1962).

METHOD

Subjects

Subjects used in the present study were 192 undergraduate students enrolled in the introductory psychology course at the University of Alberta, Edmonton, during the 1964-65 academic year. Half of the Ss were men and the other half were women. Students in the course were required to serve as Ss in a limited number of experiments during the school year as a course requirement, but participation in any particular experiment was not required. All Ss volunteered and participated in the present experiment during the first six weeks of the academic year.

Subjects were assigned to experimental conditions randomly by pairs of the same sex. Each pair was used only once for an experimental session lasting approximately 20 to 30 minutes. Data on all Ss were used except in cases of apparatus failure, or if an S gave clear evidence of not understanding the instructions. In such cases the S pair was replaced.

Instructions and Task

Instructions were presented to Ss by means of a tape recorder. Subjects were instructed that they would see a series of 96 "slides" presented on a screen; that they were to place each into one of two categories, designated by two buttons on a panel in front of them; that some of the slides correctly belonged in the "plus" category, designated by the button marked by the "plus" sign, and all others

belonged in the "unmarked" category; that a light would go on over the correct button to let them know whether their response was correct or not; that they should make as many correct responses as possible; and, that they should try to "figure out what it is about some of the 'slides' that makes them go in the 'plus' category", i.e., try to identify the concept, so that they might make all correct responses thereafter. Complete instructions are presented in Appendix I.

The task confronting each S was thus to classify each instance as it appeared on the screen into one of two categories by pressing one of the two buttons.

Variables and Design

The variables of major experimental interest were concept size, or number of relevant binary dimensions contained in a problem, and proportions of positive and negative instances. Three levels of concept size were used; one, three, and five relevant nonredundant visual dimensions. Four levels of proportional divisions of positive and negative instances were employed, as follows: 1/8 positive and 7/8 negative; 2/8 positive and 6/8 negative; 3/8 positive and 5/8 negative; and 4/8 positive and 4/8 negative. Hereinafter this variable will be designated as "proportion of positive instances", or simply "proportion", and a level of the variable will be designated by the proportion of positive instances only, e.g., "3/8 positive", or "3/8".

Two additional variables were employed. One of these was sex,

since it was of secondary interest to determine any sex differences which might exist in connection with the variables of interest. The other variable was <u>problems</u>. Two different problems were used at each level of concept size to reduce the likelihood of results idiosyncratic to particular dimensions, levels of dimensions, or combinations. A problem here is defined as a particular combination of dimensions and levels of dimensions. Since the number of dimensions determines concept size, obviously different problems had to be employed at each level of concept size. As such, problems were nested within each level of concept size, constituting a partially heirarchical design. All other variables were orthogonal. One irrelevant dimension was used in each problem to elevate problem complexity.

Since each problem contained 96 instances, it was possible to divide these into six <u>blocks</u> of sixteen instances each for repeated measures analysis. This required that the proportional division hold for each block separately in order to make them comparable.

The design was thus a 3 x 4 x 2 x 2 x 6, partially orthogonal, partially heirarchical design, with repeated measures. Subject pairs were randomly assigned to one of the 48 conditions, with two pairs falling in each cell for a total of four <u>S</u>s per cell. A given <u>S</u> or pair of <u>S</u>s received only one problem.

<u>Apparatus</u>

A single experimental room was used for running all <u>S</u>s. The room was divided into two portions by a fiberboard partition, in the middle of which was a translucent 8 x 11 in. screen surrounded by a

$1\frac{1}{2}$ in. black border. The two Ss sat facing the screen at small individual tables equipped with response-feedback panels. The S's edge of the table was $49\frac{1}{2}$ in. from the screen. Between the two Ss was a movable partition 52 in. high and 42 in. wide which shielded Ss from each other and from each other's response-feedback panel. Lighting was very dim during the experimental procedure so as to allow clear perception of the stimulus configurations appearing in the screen, yet enough light was available so that Ss could clearly see their response-feedback panels.

The response-feedback panels were made of wood and painted grey. Two response buttons, 6 in. apart, were set in the slanted response panel 13 in. from S's edge of the table. Behind these, on a vertical board projecting above the button panel, were two green lights, aligned vertically with the response buttons. Beneath one of the two buttons was a large "plus" sign made of black tape. The "plus" button was on the right on one of the two panels and on the left for the other.

Each time a S pressed one of the two response buttons the light over the correct button went on. It remained on until two seconds after the last of the two Ss responded, when both Ss' lights went out.

Behind the partition was located the remainder of the apparatus, including a 16 mm. Dunning-Animatic strip film projector fitted with a Bell and Howell, one in., F/1.9, Increlite projector lens. The lens projected the stimuli onto the screen through an area surrounded by a cardboard rectangular enclosure to reduce the amount of stray light on the screen so as to maintain relatively constant

lighting conditions and maximal resulution of the stimulus figures.

Two Lehigh Electronic event recorders, one for each S, were used to record Ss' responses as well as the type of instance--positive or negative--presented. The latter were recorded automatically on the same tape record for ease of scoring and as a check on the synchronization of program tape and strip film. Total time taken per pair of Ss was recorded by E, using a stopwatch, and included the time from onset of the first instance to the onset of the first blank frame following the 96 instances.

The entire system was automatic, requiring only turning the system on and then off at the end. Electrical components and controls were built into a panel for ease of operation. This included a light sensitive relay mechanism activated by a punch tape which was synchronized with the strip film. For each of the 24 film strips there was a corresponding program tape. This tape determined which of each S's two lights would go on when he responded to each of the 96 instances. The light corresponding to the correct response button went on immediately following each response and remained on until the last of the two Ss responded. A time delay relay then was activated, and the light corresponding to the correct response button on each S's panel, as well as the strip film frame, remained on. At the end of the two second delay, the feedback lights went out and both the strip film and program tape advanced simultaneously, marking the onset of a new trial or instance.

Prior to starting the experiment all combinations of types of stimuli and responses were checked so as to eliminate the possibility

of differential cues being presented to one or both Ss as a result of such things as the other S's feedback lights being reflected off the walls, or from differentially clacking relays, or differences in the movement of the recorder pens. All such cues were removed. An S could tell when his partner responded, but not which response he made or which feedback light went on.

Problems, Programs, and Stimuli

Each problem consisted of a series of 96 instances, or stimulus configurations. These were composed of combinations of up to six binary visual dimensions, as follows: (1) shape, squares or circles; (2) size, large or small; (3) number, two or three; (4) vertical position, top or bottom; (5) horizontal position, right or left; and (6) orientation, vertical or horizontal array of figures.

A positive instance of the concept was defined as a particular level of a relevant dimension or combinations of positive levels of relevant dimensions.

Since the three levels of concept size consisted of one, three, and five nonredundant binary dimensions, in addition to one irrelevant dimension in each problem, this required that two, four, or six dimensions vary in a given problem. Those dimensions not used in a problem were held constant at a randomly determined level.

Differences between dimensions used, and differences between problems, were not of particular experimental interest in the present study. Archer (1962) has pointed out, however, that visual dimensions vary in terms of how discriminable and salient they are. He refers

to this feature as "obviousness". He stated that the positional dimensions are particularly low in obviousness. The present study includes two such dimensions, horizontal and vertical position. In an effort to maximize relative homogeniety of difficulty levels of problems, both of these dimensions were used in each problem at each level of size. Either horizontal or vertical was included as the irrelevant dimension in each problem, while the other of these was one of the relevant dimensions. Thus, when horizontal position was a relevant dimension, vertical position was the irrelevant dimension, and vice versa. At the three-relevant-dimension size, two dimensions were selected randomly from the remaining four dimensions, and these were included in both problems at that level. At the five-relevant-dimension size, the remaining two dimensions were added. Positive levels of relevant dimensions were determined by chance for each of the six problems separately. Table 1 shows the dimensions used in each problem, their positive relevant levels, and constant levels of unvarying dimensions.

Four separate programs were required for each problem, one for each level of proportion of positive instances. The proportional division of positive and negative instances was maintained in each of the six blocks of 16 instances. Separate random sequences were constructed for each of the 24 separate programs. Because blocks of instances were treated as equal units for purposes of repeated measures analysis, the positive instances were assigned to random positions within each block of 16 instances. Since there were two possible types of positive instances due to the inclusion of one irrele-

Table 1

Problems at Each Level of Size

Size	Problem	Relevant Dimensions : Positive Relevant Level	Irrelevant Dimension	Constant Dimensions : Level
1	A	Horizontal : Right	Vertical	Number : Two Size : Small Shape : Circle Orientation: Horizontal
	B	Vertical : Bottom	Horizontal	Number : Two Size : Large Shape : Circle Orientation: Horizontal
3	C	Number : Three Size : Small Horizontal : Right	Vertical	Shape : Square Orientation: Horizontal
	D	Number : Two Size : Small Vertical : Top	Horizontal	Shape : Circle Orientation: Vertical
5	E	Number : Three Size : Small Horizontal : Right Shape : Square Orientation: Horizontal	Vertical	
	F	Number : Two Size : Large Vertical : Top Shape : Circle Orientation: Vertical	Horizontal	

vant dimension, each of the two types appeared once, twice, three, or four times in a block for the 1/8, 2/8, 3/8, and 4/8 proportions, respectively. The only restriction on the random assignment of positive instances was that they not all be adjacent to each other within a block. Negative instances were then placed in the remaining positions. The various types of negative instances were also distributed randomly.

Each stimulus configuration was first drawn in black ink on a sheet of white paper, and then mounted on thin cardboard. The configurations consisted of large or small (large were twice the diameter or width of small) circles or squares, in arrays of two or three, in one of the four quadrants of the paper. Each of the stimulus cards also contained a cross in the center. This was a constant appearing in all frames. It was used mainly to provide positional reference, since the background in the frames when projected was dark against a dark surrounding.

The stimulus cards were then photographed one frame at a time, according to the programs, with a 16 mm. Bolex reflex movie camera with a zoom lens. High contrast black and white film was used for maximum definition.

The projected sizes of the individual circles and squares were approximately 1 in. and $\frac{1}{2}$ in. across for the large and small respectively.

Procedure

Subjects were seated at the two tables, and the partition was positioned between them. It was then explained to Ss that they would

first be played the tape recorded instructions, following which they would have an opportunity to ask questions before the experiment began. It was requested that the two Ss not communicate with each other from this point on until they were finished with the experiment.

The tape recorded instructions were then played. Following the instructions, Ss were asked if they had any questions. Questions were always answered by E with a paraphrase or repeat of parts of the instructions. Usually a question was asked as to how to classify the first slide. This was always answered, "on the first slide you will have to guess". In cases where this question did not arise, E volunteered this information. In most cases only a few questions arose. When both Ss indicated that they understood the task, the lights were turned out, and following a delay of at least a minute to permit adjustment to the reduced light intensity level, the switch was turned on which set the system into automatic operation, including the presentation of the first instance.

When the entire series of 96 instances was completed, the lights were turned on and Ss were asked to fill in a questionnaire asking them, in addition to more general questions, to describe the positive instances, and asking them if they pressed one button more frequently than the other prior to identifying the concept, and why. This questionnaire is shown in Appendix II.

At the end of the session, Ss were asked if they were able to "figure out what went into the 'plus' category" to be certain that instructions had been understood and followed. In all but one case, which was discarded and replaced, there was no evidence that Ss had not understood the instructions. Generally Ss found the task to be enjoyable.

RESULTS

Error Analyses

The primary measure employed in the present study was errors per block of sixteen instances. Differences between levels of the variables and interactions were tested for significance by analysis of variance. The results are presented in Table 2.

Analysis of the between Ss variance indicated that the size and proportion main effects were both significant. The interaction between these two variables failed to reach an acceptable level of statistical significance, as is the case with the remaining sources of between Ss variance. Figure 1 shows mean errors as a function of size for each proportion of positive instances. The 4/8 group made the fewest total errors, followed by the 1/8 group, then the 2/8 group, and the greatest number of errors was made by the 3/8 group. It is also apparent from inspection of Figure 1 that mean errors increased as a function of concept size.

Trend tests (Winer, 1962) on these variables were carried out as shown in Table 3. On the proportion variable the quadratic and cubic components of the trend were significant, but the linear component was not. This indicates that mean errors as a function of proportion of positive instances did not vary significantly in overall slope from a horizontal straight line, but that the function was significantly curvilinear about a straight line with zero slope. The trend test on the size variable indicated that both the linear and quadratic components were significant. Thus, mean errors varied as

Table 2

Analysis of Variance of Total Errors

Source	df	MS	F
Size	2	752.95	38.20**
Problems/Size	3	7.07	
Sex	1	23.06	1.17
Proportion	3	197.95	10.04**
Size x Proportion	6	21.24	1.08
Size x Sex	2	14.15	
Sex x Problems/Size	3	7.05	
Proportion x Problems/Size	9	23.95	1.22
Sex x Proportion	3	40.22	2.04
Size x Sex x Proportion	6	7.26	
Sex x Proportion x Problems/Size	9	6.79	
Ss/Groups	144	19.71	
Blocks	5	543.63	210.63**
Blocks x Size	10	18.85	7.30**
Blocks x Problems/Size	15	3.20	1.24
Blocks x Sex	5	2.38	
Blocks x Proportion	15	6.90	2.67**
Blocks x Size x Sex	10	1.33	
Blocks x Size x Proportion	30	2.88	1.11
Blocks x Sex x Problems/Size	15	.63	
Blocks x Proportion x Problems/Size	45	4.55	1.76*
Blocks x Sex x Proportion	15	.71	
Blocks x Size x Sex x Proportion	30	2.64	1.02
Blocks x Sex x Proportion x Problems/Size	45	1.85	
Blocks x Ss/Groups	720	2.58	
Total	1151		

* $p < .01$
** $p < .001$

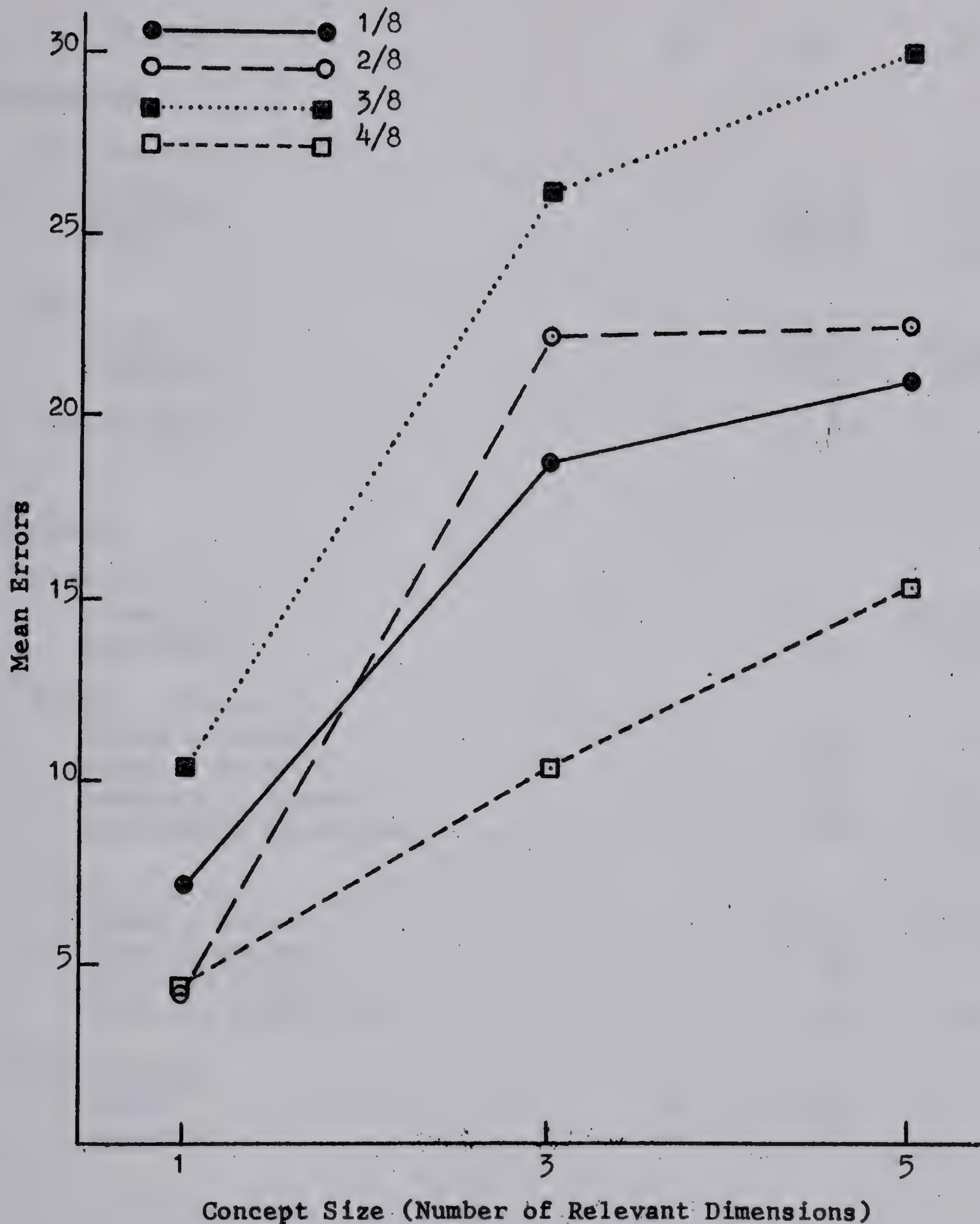

Fig. 1. Mean errors as a function of concept size for each level of proportion of positive instances.

Table 3

Orthogonal Polynomial Trend Analyses on Major Variables

Source	df	MS	F
Between Ss			
Proportion			
Linear	1	49.69	2.52
Quadratic	1	328.32	16.66**
Cubic	1	215.84	10.95*
Size			
Linear	1	1330.89	67.52**
Quadratic	1	175.01	8.88*
Between Error	144	19.71	
Within Ss			
Blocks			
Linear	1	2276.19	504.03**
Quadratic	1	391.57	119.49**
Blocks x Proportion			
Linear x Linear	1	31.68	7.01*
Linear x Quadratic	1	.06	
Quadratic x Linear	1	3.93	1.20
Quadratic x Quadratic	1	3.48	1.06
Blocks x Size			
Linear x Linear	1	109.17	24.17**
Linear x Quadratic	1	52.59	11.65**
Quadratic x Linear	1	.41	
Quadratic x Quadratic	1	10.87	3.32
Within Error			
Linear	144	4.52	
Quadratic	144	3.28	

* $p < .01$

** $p < .001$

a negatively accelerated function of number of relevant dimensions for the conditions of one, three, and five relevant dimensions.

Despite the fact that the overall interaction was not significant, interactions between the trend components of the size x proportion interaction term were calculated through quadratic x quadratic and none of the four trend combinations was found to reach significance.

It should be noted that none of the between Ss sources of variance that were not of primary experimental interest achieved statistical significance. This includes the sex variable and pooled problems nested within sizes, as well as all interactions involving these variables. Therefore, performance between the two sexes did not differ overall, and performance means on the two problems contained in each size were also homogeneous.

Inspection of the within Ss variance F-ratios in Table 2 shows that the repeated measures variable of blocks was significant, as might be expected, and the interactions of blocks with size and with proportion were also significant. In addition, the blocks x proportion x problems/size reached significance.

To test the significance of trend components and interactions of trend components in the within Ss sources of variance the appropriate components of the blocks x Ss/groups were calculated and used as error terms. These are presented in Table 3. Since there is some controversy regarding the appropriateness of using the pooled error variance as opposed to these components (Gaito and Wiley, 1965), the trend variances were also tested with the pooled error mean square.

The results are consistent in that all F-ratios either did not reach significance or were significant at $p < .01$ or better with both error terms, with the exception of the blocks-quadratic x size-quadratic interaction. Testing of this interaction by the component error variance yielded no significance, but with the pooled error variance it reached the $p < .05$ level. Gaito and Wiley compared the two types of error estimates and made a strong case for the use of the appropriate component error terms. As such, these are reported in Table 3.

Figure 2 shows a negatively decelerating error trend as a function of blocks of instances for all groups combined. An initial analysis of trend components indicated that all components were significant through quartic using the component error terms (Grant, 1956), and through cubic using the blocks x Ss/groups term, even though inspection of the curves in Figure 2 suggests nothing higher than a quadratic component. Grant cautions that such unexpected statistical significance may occur in the case of logarithmic and exponential functions when the data are highly reliable. Therefore, the validity of the trend test is questionable in this particular case. The linear and quadratic Fs are presented nonetheless in Table 3. Fitted values were obtained by the orthogonal polynomial method (Winer, 1962) as shown in Figure 2, and it is obvious that the form of the best fitting curve is a decreasing logarithmic function.

The blocks x size interaction was significant. This was probably due in part to the fact that the one-relevant-dimension group began to approach zero errors relatively early in the series of blocks, and was thus restricted by a natural limit. Both the blocks-linear x

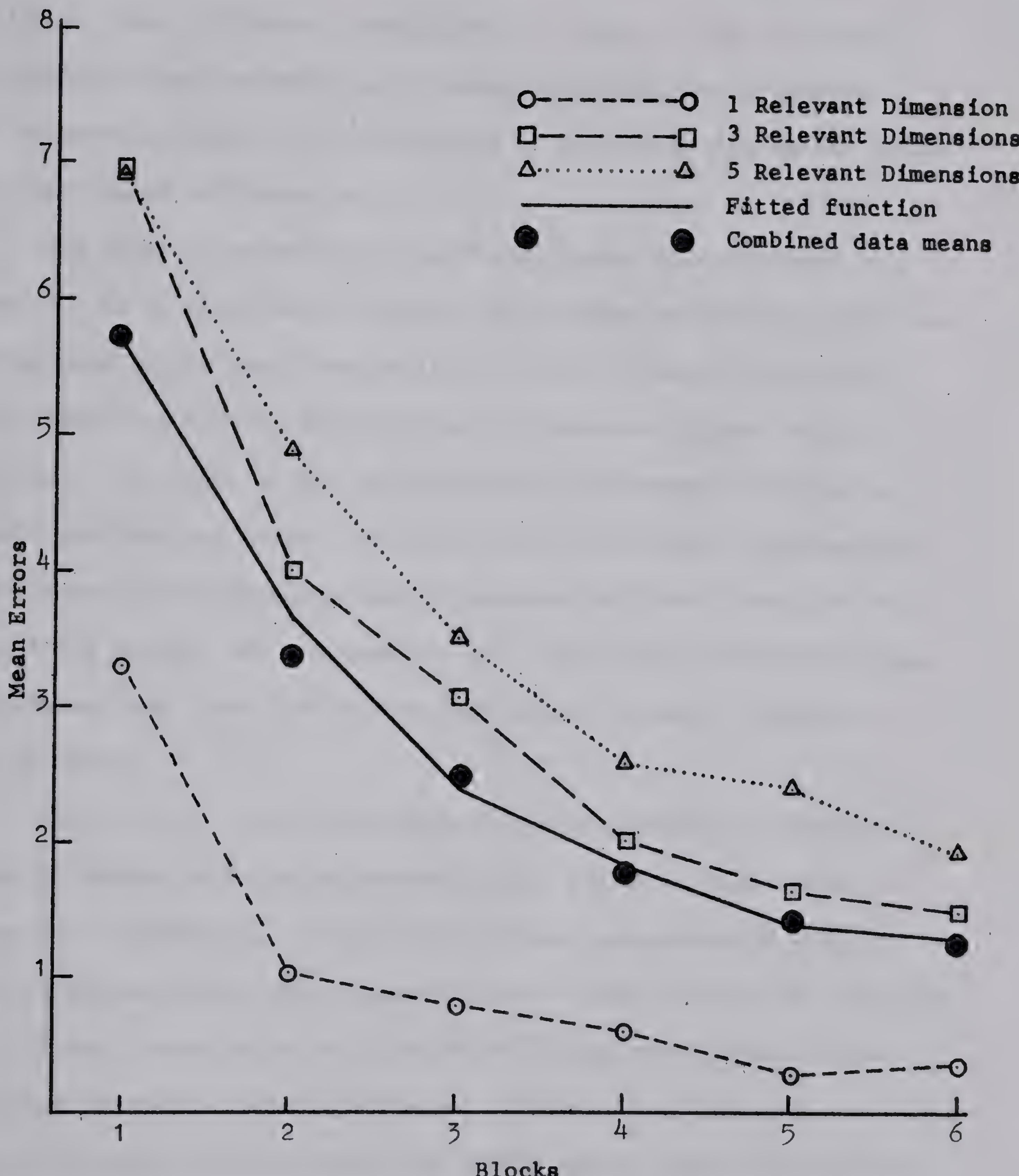

Fig. 2. Mean errors as a function of blocks of 16 instances for each level of size. Also shown is the fitted function for the combined levels of size. The black circles show the combined size means in relation to the fitted function.

size-linear, and blocks-linear x size-quadratic components were significant (Table 3). The first of these indicated that the slopes of the trends for the levels of size, plotted against trials, were not parallel. From a different vantage point, slopes of the six trends for trials plotted as functions of size approached the horizontal with succeeding trials. The curvatures of the three size curves varied as a function of different trials.

The blocks x proportion significant interaction is shown in Figure 3. It is clear that the order of the four proportion conditions was the same on the last four trials as it was on mean total errors. In the first two blocks, however, the 1/8 means are higher than the 2/8 means. The order of the two functions then becomes reversed between blocks two and three. It appears from the curves that the rate of performance improvement abruptly decreased in block three for the 2/8 and 3/8 groups, but continued on for the 1/8 and 4/8 groups through block three, and block four showed a relatively greater decrease in rate of change.

Tests on the trend interactions through quadratic x quadratic (Table 3) showed that the proportion-linear x blocks-linear interaction was significant. A plot of the linear components of each of these variables against the other would show that they are not parallel.

Table 2 shows also that the pooled blocks x proportion x problems/size interaction was significant. Inspection of the data revealed that most of the variance was contributed by the five-relevant-dimension level of size, followed by the three-dimension level, and then the one-dimension level. The shapes of the error curves repre-

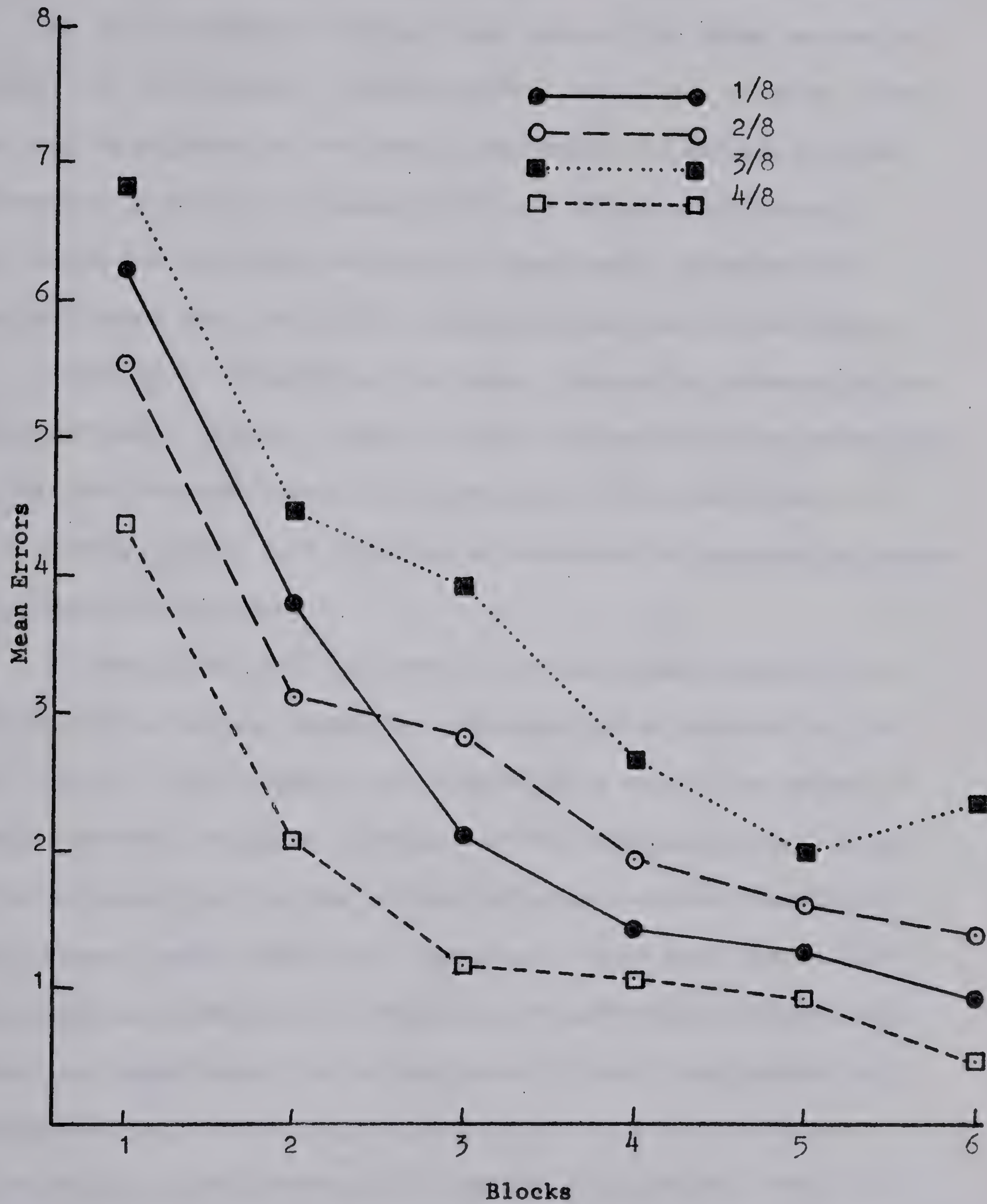

Fig. 3. Mean errors as a function of blocks of 16 instances for each level of proportions of positive instances.

senting proportions for each of the six problems indicated greater differences between proportions and between problems as size was increased. Generally, the functions deviated from parallel considerably.

It will be noted in Table 2 that none of the other sources of variance was significant. Problems within each level of size, therefore, may be regarded as relatively homogeneous in effect, and the performance by men and by women Ss did not differ significantly. Since these two variables were not of experimental interest their nonsignificance does not affect the primary purposes of the study.

Contrary to expectation, the size x proportion interaction was not significant. Figure 1 shows a slight suggestion of an interaction, but this did not even approach significance. Thus, performance of the proportion groups as a function of size must be regarded as essentially parallel functions.

It should be noted that error score means were somewhat correlated with variances, producing heterogeniety of variance in the error scores. This situation was created by a ceiling on number of possible correct responses. Because of the nature of concept learning tasks, once S solves the problem his errors should immediately drop to zero, beyond which they cannot go. Since this occurs differentially as a function of conditions if differences between conditions are significant, as in the present study, this results in heterogeniety of within cells variance. This state of affairs is not regarded as invalidating as is pointed out by Winer (1962), in the light of a growing body of literature on the topic. As a safeguard, Lindquist (1953) recommends using a more stringent rejection

level than would otherwise be used. Since all sources of variance were significant at $p < .01$ or better in the present analysis of errors, the results should be regarded as valid.

Concept Attainment

In the assessment of concept identification performance, errors per block constitutes only one of the dependent variable measures possible. Another way of looking at the present data is in terms of the number of Ss attaining the concept as a function of the variables of interest. Analysis in terms of frequencies of attainment yields a type of assessment which may or may not be correlated with total errors.

A. Criterion: Since for a given problem, blocks were equal in terms of number of positive and negative instances, it was possible to establish a *post hoc* criterion of concept attainment in terms of errors per block. Two such criteria were considered: one block with no errors, and two blocks with no errors, in both cases permitting a maximum of one error per block following the qualifying block(s). Permission of one error per block following attainment without disqualification was included because of the occasional report by Ss that they made "careless" errors after solving the problem. Indeed, this was borne out in that there was a negative relationship between concept size and the number of such errors following attainment of either of these two criteria. That is, the simpler the problem the greater was the number of Ss attaining the concept, and the greater was the number making a single error following attainment,

by either criterion. In only two cases did an S make two errors in different blocks following the perfect block(s). None made more. On this basis then, the occasional errors following attainment were not disqualifying. As for the definition of attainment as one or two perfect blocks, one block appeared to be the more reasonable since under the two block criterion 19 Ss were able to verbalize the concept correctly without having attained the criterion. When the one block criterion was employed, however, only nine of these "false negatives" remained. This number could more easily account for those Ss who attained the concept during the last block after making some errors, and therefore this appeared more valid as a criterion of attainment.

B. Attainment: The overall number of attainers by this criterion was 128, or 67% of the Ss. (By the more stringent criterion 106 Ss, or 55.2%, would be attainers.) Among these, the modal block of attainment was block 2 for the one-relevant-dimension group, and block 4 for each of the three- and five-relevant-dimension groups. Table 4 shows the distribution of attainers and nonattainers by concept size. It is clear that the distribution of attainers was consistent with the error results in that the proportion of attainers decreased as size increased. This distribution was tested by chi-square and found to be statistically significant ($\chi^2 = 41.39$, 2 df, $p < .001$). Frequencies of attainment vs. nonattainment by proportion of positive instances as shown in Table 5 were also found to be significant ($\chi^2 = 10.88$, 3 df, $p < .02$), and the distribution of attainers again was consistent with the error results in that the 4/8 condition had the highest proportion of attainers followed by 1/8, 2/8, and 3/8

Table 4

Frequency Distribution of Attainers-Nonattainers by Size

	Size 1	Size 3	Size 5
Attainers	61	40	27
Nonattainers	3	24	37

$\chi^2 = 41.39$; $df = 2$

$p < .001$

Table 5

Frequency Distribution of Attainers-Nonattainers by Proportion

	Proportion 1/8	Proportion 2/8	Proportion 3/8	Proportion 4/8
Attainers	35	33	23	37
Nonattainers	13	15	25	11

$\chi^2 = 10.88$; $df = 3$

$p < .02$

in that order. It is of interest to note that exactly half the attainers were men and half were women.

C. <u>Concept Verbalization</u>: One of the items on the Subject Information Questionnaire was: "Describe the slides that went into the 'plus' category". While such verbalization of the concept did not constitute primary data in the present study, these data were nonetheless analyzed since concept learning studies have often used concept verbalization as their primary measure of concept learning. The results might then be more easily compared with such studies by the interested reader.

Subjects were categorized as verbalizers if they stated all positive values of all relevant dimensions, and if they were accurate when mentioning irrelevant or constant dimensions. Table 6 shows the distribution of verbalizers-nonverbalizers by attainers-nonattainers. While 76.6% of the attainers were able to accurately state the concept in writing, it is of interest to note that nine nonattainers, or 14.1% of the nonattainers, were also able to do this. This suggests that some may have attained the concept during the last block.

The 128 attainers were categorized by verbalization and concept size in Table 7. The interaction chi-square was not significant ($\chi^2 = 5.72$, 2 df, $p < .10$), but does approach an acceptable level. The same was true for the verbalization by proportion frequency distribution as shown in Table 8 ($\chi^2 = 5.81$, 3 df, $p < .20$). The percentage of attainers able to verbalize the concept was less for the 4/8 condition than for any of the others, even though the percentage of attainers was the greatest for this condition. Only 64.9% of attainers

Table 6

Frequency Distribution of Verbalizers-Nonverbalizers by Attainers-Nonattainers

	Attainers	Nonattainers
Verbalizers	98	9
Nonverbalizers	30	55

Table 7

Frequency Distribution of Verbalizers-Nonverbalizers by Size

	Size 1	Size 3	Size 5
Verbalizers	52	29	17
Nonverbalizers	9	11	10

$\chi^2 = 5.72$; $\underline{df} = 2$

$\underline{p} < .10$

Table 8

Frequency Distribution of Verbalizers-Nonverbalizers by Proportion

	proportion 1/8	2/8	3/8	4/8
Verbalizers	31	26	17	24
Nonverbalizers	4	7	6	13

$\chi^2 = 5.81$; $\underline{df} = 3$

$\underline{p} < .20$

were able to verbalize the concept at 4/8 as opposed to 88.6% at 1/8, 78.8% at 2/8, and 73.9% at 3/8.

Types of Instances

Differential performance on positive and negative instances was of interest in an attempt to shed light on the concept learning process. However, because the numbers of positive and negative instances presented to Ss was varied, direct test of error differences on all the data was neither meaningful nor appropriate since the possibilities for error were not constant.

A. Percent Errors: For exposition purposes, percent errors were plotted. Figure 4 shows percent total errors as a function of proportion for positive and negative instances separately. Each point on each curve represents the percentage of errors out of the total possible errors at that point for that type of instance. Percentage of errors on positive instances generally decreased as the proportion or number of such instances was increased. On negative instances, however, the percentage of errors was close to 15% at all proportion levels with the exception of the 3/8 condition which had 25.2% errors. These functions showed an interesting reversal. The percentage of errors on positive instances was greater than on negative instances when their frequency was low--or when there was a great disparity between the proportions--and this relationship was reversed as the proportion of positive instances was increased and the disparity in frequencies was reduced. Looking at these curves from another vantage point, the negative instance function might be lowered at 1/8 and 2/8

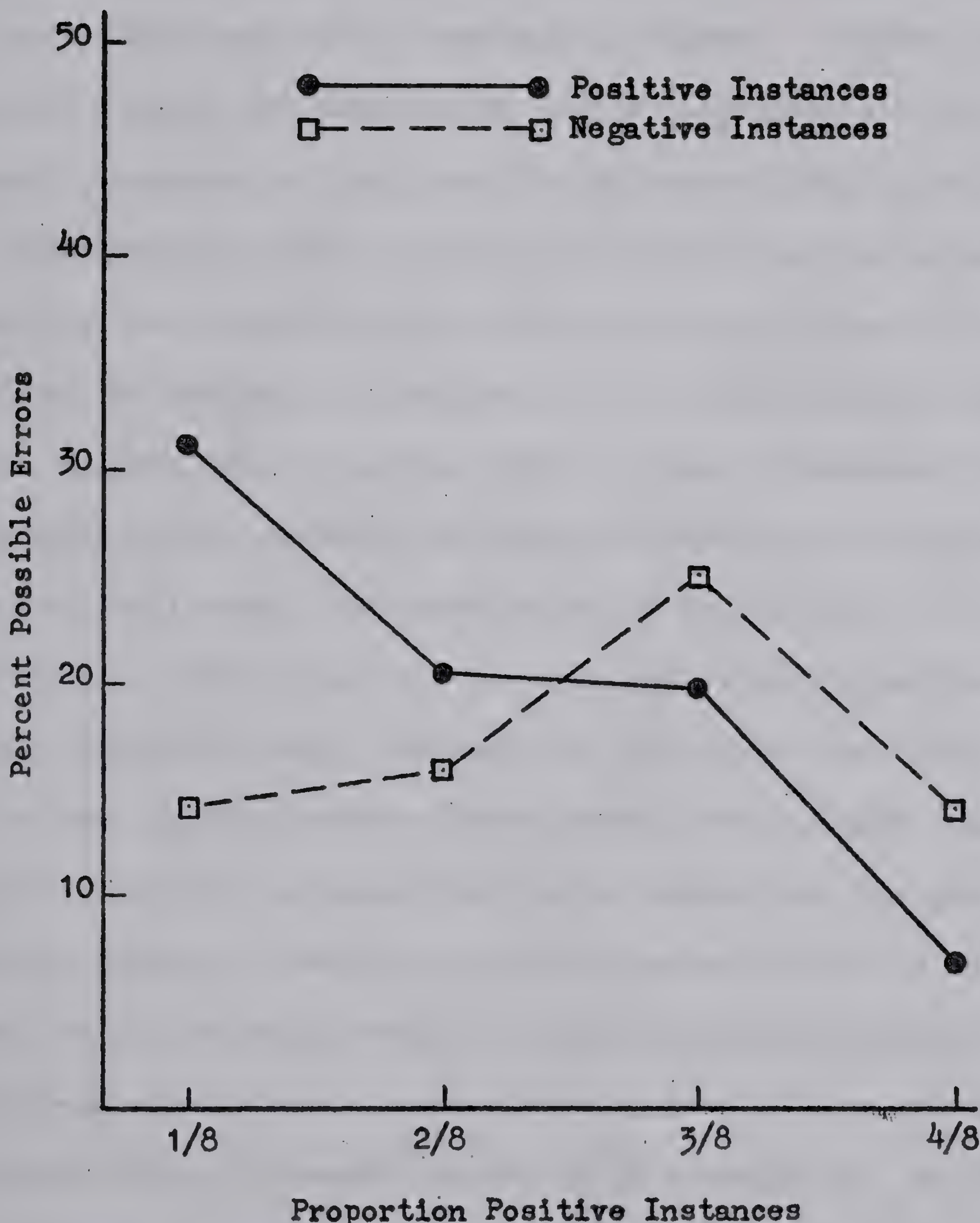

Fig. 4. Percent possible errors as a function of proportion of positive instances for both positive and negative instances.

due to probability learning in these conditions, causing the two curves to deviate from parallel.

Figure 5 shows these same data over the six blocks of trials. The functions appear to be relatively consistent from block to block in all proportion conditions. The two curves in each proportion condition are, for the most part, relatively parallel. At 1/8 the percent errors on negative instances dropped as a function of trials in a very smooth negatively decelerating manner. Positive instance errors started out considerably higher (approximately 50%, or chance level), remained at this level in the second block, then dropped in a relatively parallel fashion with negative instance errors, maintaining the greatest disparity between the two types of instances of any of the proportion conditions. In the 2/8 condition both curves are somewhat closer together with the rate of decrease of negative instance errors somewhat lessened and the slope of positive instance errors considerably decreased overall. In contrast with the 1/8 condition, percent positive instance errors on the earlier blocks were strikingly fewer. Between the 2/8 and 3/8 conditions the positive and negative instance error curves reverse, such that at 3/8 the overall negative instance function is higher than the positive instance function. While the positive instance curve is similar to that at 2/8--showing perhaps a slightly sharper slope--the negative instance curve shows a large elevation over 2/8, accounting for the greater number of overall errors in this condition. At 4/8 this relationship between the two curves is maintained, but both are lower in percent errors.

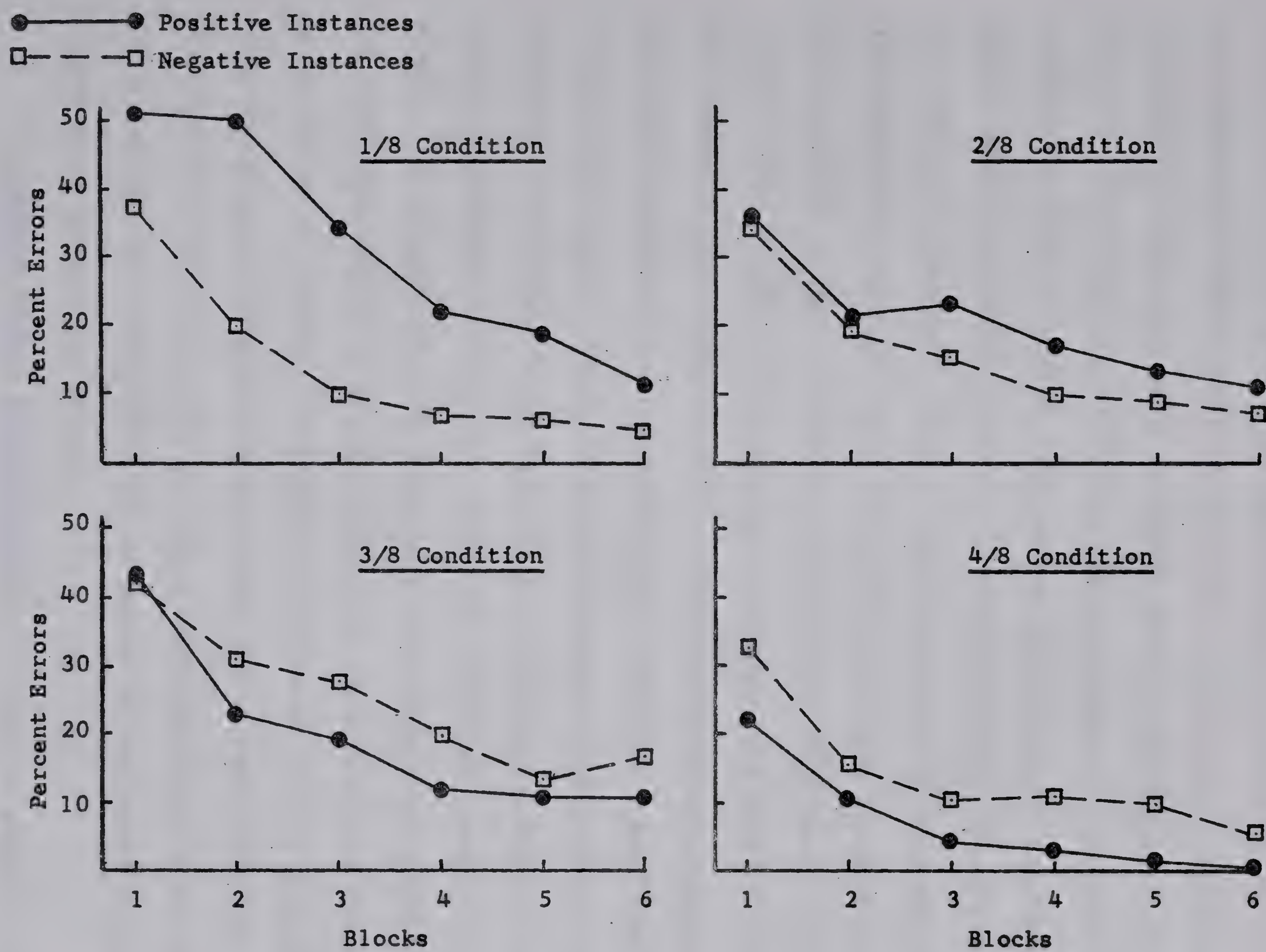

Fig. 5. Percent errors made on positive and negative instances as a function of blocks in each proportion condition.

B. Analysis of 4/8 Data: Since at the 4/8 level the proportions of positive and negative instances were equal, and thus, the possibility of error scores the same for each, it was possible to test the difference between types of instances, as well as the interaction of types with the other major variables with the exception of proportion. The 48 Ss at the 4/8 level were analyzed in a split-plot design with size constituting the between Ss variable of interest, and blocks and type instance, as well as the interactions, constituting the within Ss variables. Table 9 shows the results of this analysis. The significant size and blocks main effects and size x blocks interaction yielded no new information except in that these results with equal numbers of positive and negative instances were consistent with the results of the overall analysis. The significant instance type main effect confirmed the indication given in Figures 4 and 5 that the overall differences between positive and negative instances was significant. The nonsignificant blocks x type interaction supports the earlier observation that the two curves are relatively parallel over blocks. The significant size x instance type interaction showed that the difference between positive and negative instance errors increased as the number of relevant dimensions was increased from one through five, the greatest difference being between the one-and three-relevant-dimension conditions. At the one-relevant-dimension level the difference was very slight due to a low number of mean errors--close to zero--over most of the blocks.

C. Analyses Within Instance Types: In order to compare positive instance performance as a function of the (approximate) number

Table 9

Analysis of Variance of Errors at 4/8 Proportion Condition
Including Type of Instance as a Variable

Source	df	MS	F
Size	2	40.89	11.09**
$Error_a$	45	3.69	
Blocks	5	48.06	64.43**
Type	1	50.77	68.05**
Blocks x Type	5	.80	1.08
Size x Blocks	10	2.03	2.72*
Size x Type	2	13.63	18.27**
Size x Blocks x Type	10	1.21	1.63
$Error_b$	495	.75	
Total	575		

* $p < .01$
** $p < .001$

of negative instances randomly interspersed, i.e., proportion of negative instances, an equal number of positive instances was analyzed for all proportion conditions regardless of number of intervening negative instances. That is, mean errors on the first twelve positive instances were compared for each proportion condition by size. (All proportion conditions had twelve or more positive instances.) The same procedure was followed to obtain error scores on the first 48 negative instances. (All proportion conditions had at least 48 negative instances.)

The results of analyses of variance of error scores are presented in Tables 10 and 11 for positive and negative instances, respectively. Both the proportion and size main effects were significant in each case, but the interactions were not. Figures 6 and 7 show errors as a function of size for each proportion condition. In Figure 6 it appears that for positive instances the 1/8, 2/8, and 3/8 curves are clustered, while the 4/8 curve is very superior at the three-relevant-dimension level, and remains superior at the five-relevant-dimension level. Except at 4/8, performance on positive instances is superior at the five-relevant-dimension level as compared with the three-relevant-dimension level.

For negative instances, Figure 7 shows the 4/8 condition superior to all others, the 3/8 condition the worst, and the 1/8 and 2/8 close together about midway between. The relationships between the four curves are very similar to those which occur in the main analysis of the overall data, presented in Figure 1.

The meaning of the proportion variable in these two analyses

Table 10

Analysis of Variance of Errors on First 12 Positive Instances

Source	df	MS	F
Proportion	3	29.59	5.99**
Size	2	143.77	29.12**
Proportion x Size	6	9.90	2.00
Ss/Groups	180	4.94	
Total	191		

** $p < .001$

Table 11

Analysis of Variance of Errors on First 48 Negative Instances

Source	df	MS	F
Proportion	3	328.08	9.74**
Size	2	1787.83	53.06**
Proportion x Size	6	20.43	
Ss/Groups	180	33.69	
Total	191		

** $p < .001$

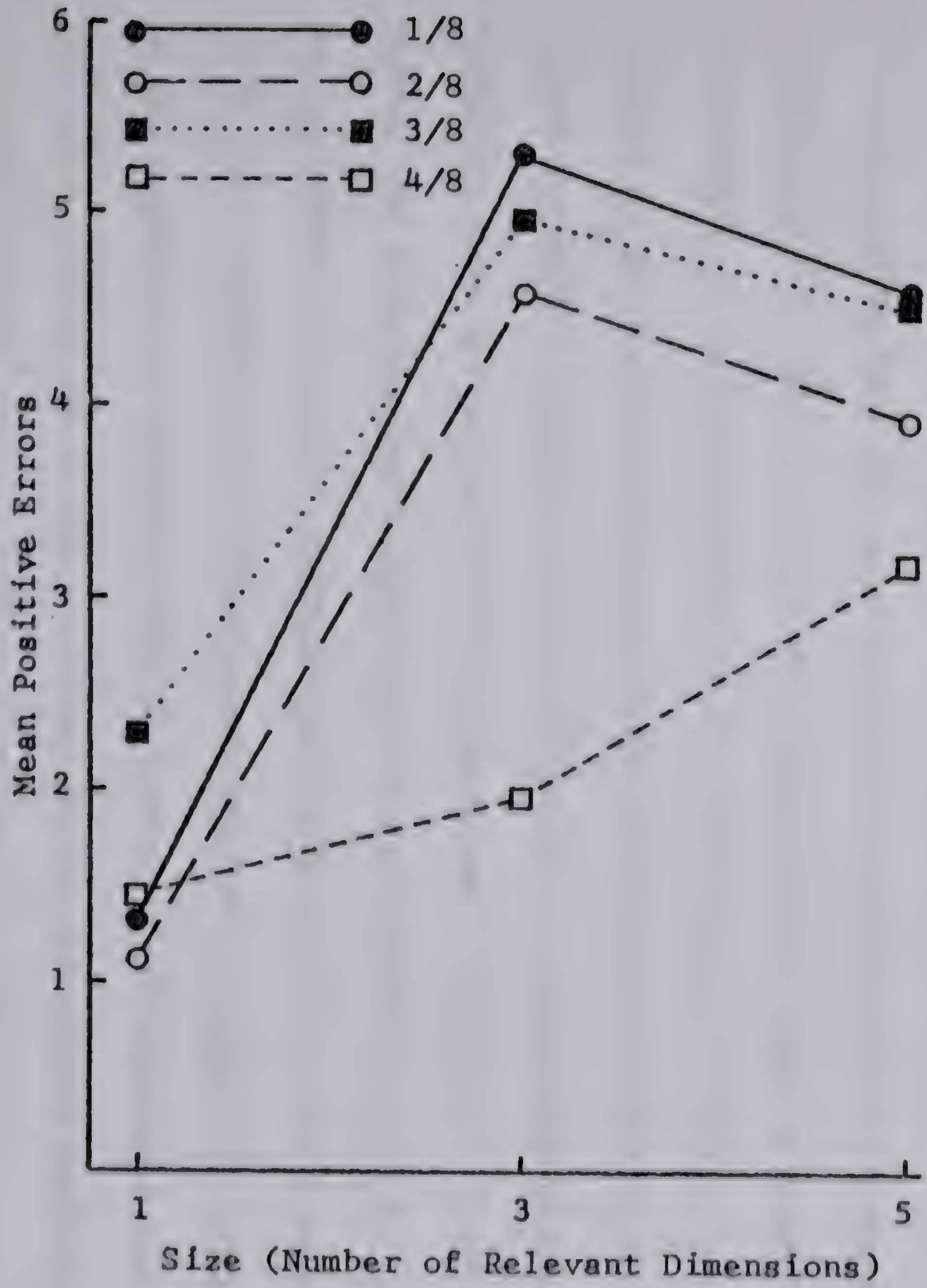

Fig. 6. Mean errors on first 12 positive instances as a function of size for each proportion condition.

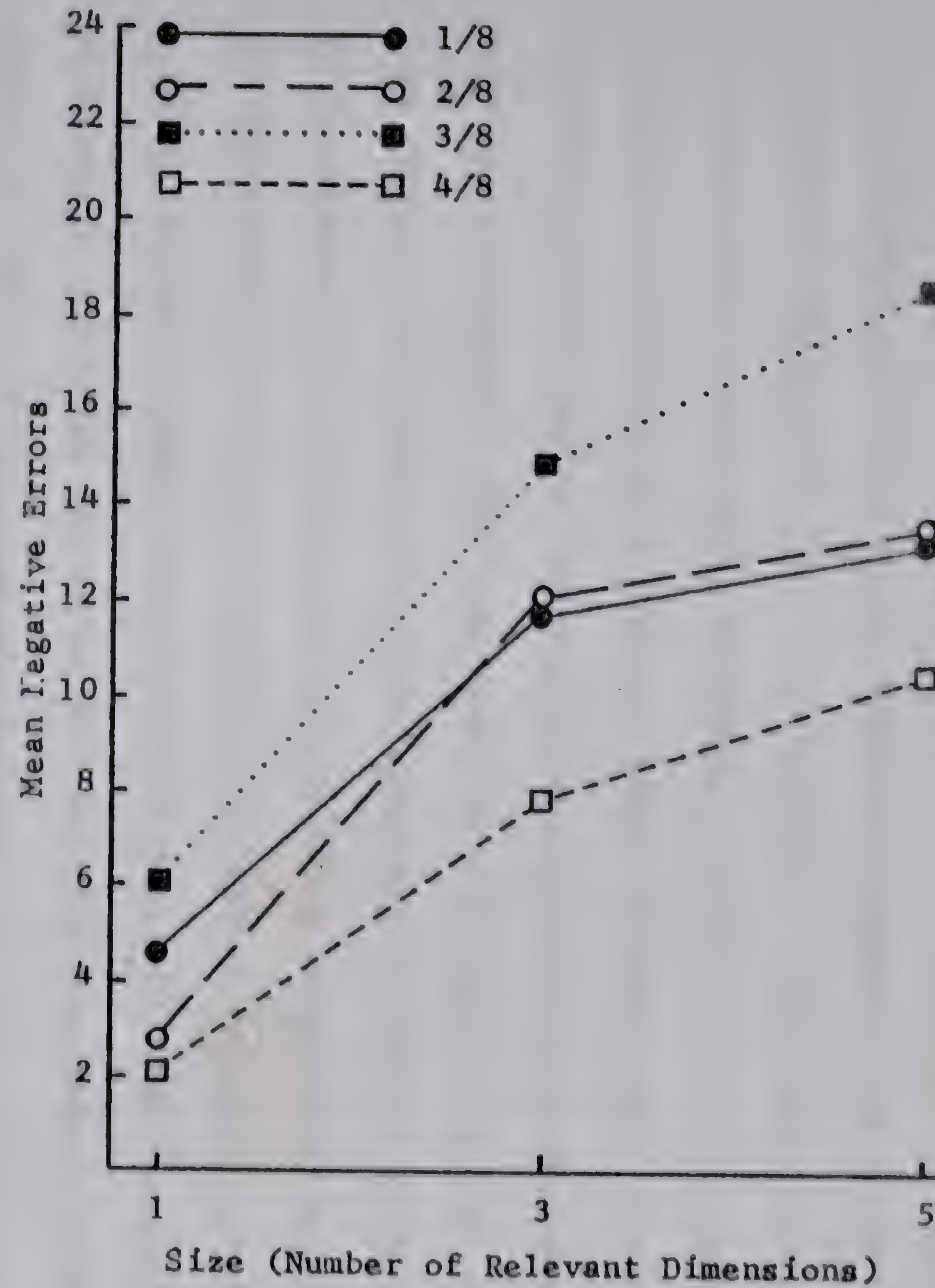

Fig. 7. Mean errors on first 48 negative instances as a function of size for each proportion condition.

is not the same as its meaning in the other analyses. In considering errors on only the first twelve positive instances, the proportion of positive instances is thus held constant. Differences between the four proportion groups must, therefore, be attributable to other factors associated with the proportion levels in this case. The obvious variables distinguishing the proportion levels here are number of interspersed negative instances together with distribution of practice. Less obvious, but possibly of importance, are systematic differences in Ss expectations of the number of remaining instances at any point, and experience in the task situation, the latter carrying with it the possibility for learning the category frequencies. For the analysis of negative instances, the proportion variable here refers to the effect of number of interspersed positive instances and distribution of practice, along with these other less obvious variables.

Assuming that the major influence on the results is the number of the opposite type of instances interspersed, Figure 6 suggests that number of negative instances has little or no facilitative influence on positive instance performance since the condition with the fewest negative instances, 4/8, is superior to the others which are higher in mean errors and clustered together. Except for the 3/8 curve, the order suggests that negative instances may have an interfering effect on positive instance performance. The differences found in Figure 7 indicate that performance was best on negative instances in the condition having the greatest number of interspersed positive instances, but worst in the condition having the second greatest number interspersed. The conditions with the least positive instances

were close together and in between. Thus, there seems to be no clear trend on the proportion variable if it is interpreted to refer only to interspersed positive instances.

If the possible benefit from probability responding and its likelihood are considered, then the order of the curves in Figure 7 lends itself to interpretation. The 1/8 and 2/8 groups stood to benefit most by probabilistic responding, and the frequency differential was probably most obvious in these conditions. At 3/8, however, obviousness was probably lower and potential gain was also minimal. Statements on the "Subject Information Questionnaire" are consistent with this reasoning. Therefore, it is possible that the 1/8 and 2/8 curves are "out of order" in the sense that they not only reflect the influence of different numbers of interspersed positive instances, but also reflect a greater tendency for Ss to respond on the basis of learned frequencies enabling them to reduce the number of errors made on negative instances. Without such benefit the 1/8 and 2/8 error curves may have been higher than the 3/8 function. The similarity between the 1/8 and 2/8 conditions might indicate that in this range the net effect of performance facilitation by interspersed positive instances and by probabilistic responding may be approximately equal.

Probabilistic responding should be reflected in positive instance performance by an increase in positive errors, this being greatest for the condition with the smallest proportion of such instances. This appears to be the case, except for the 3/8 condition, if any trend is discernible. The elevation in errors at the three-relevant-dimension level as compared with the five-relevant-dimension

level, for the 1/8, 2/8, and 3/8 groups, might well have occurred because Ss responded on the basis of differential frequencies to a greater extent at this level.

Pre-attainment Responding

A. Errors: Mean errors per block on Ss prior to attainment of the concept were plotted for each of the proportion conditions. These curves are shown in Figure 8. It is clear that pre-attainment performance was consistent with overall mean performance on the proportion variable in that the curves fell in the same general order. This suggests that the results obtained in the earlier analyses of error scores were not simply a function of differential rates of concept attainment.

B. Probabilistic Responding: In order to determine whether Ss were aware of differential frequencies of positive and negative instances, and used this information in their responding, they were asked on the Subject Information Questionnaire (Appendix II) whether they responded more frequently with one of the two buttons prior to attaining the concept, and to state the reason if their answer was affirmative. Those Ss who answered in the affirmative, and who gave the reason that they discovered a differential in frequencies between the two categories, were tallied by size and proportion condition. At the one-relevant-dimension level none of the Ss indicated awareness of the differential frequencies. At the three-relevant-dimension level, out of 16 Ss per proportion condition, the frequencies were as follows: 12 at 1/8; 4 at 2/8; 2 at 3/8; and 0 at 4/8. At the five-

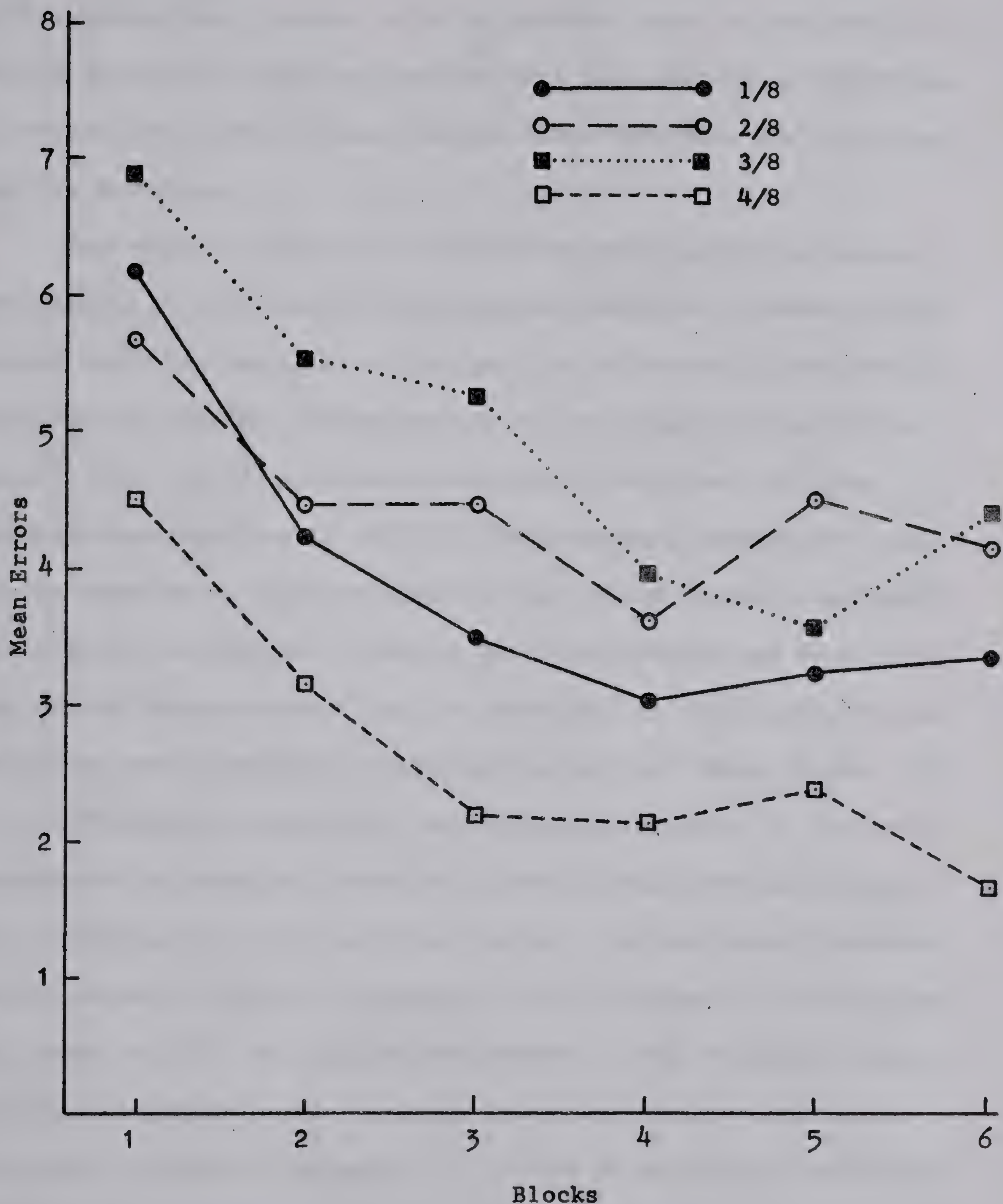

Fig. 8. Mean errors as a function of blocks of trials prior to attainment of criterion for each level of proportion.

relevant-dimension level the frequencies were as follows: 5 Ss at 1/8; 3 at 2/8; 1 at 3/8; and 2 at 4/8. It appears that the differential may have been less apparent as size was increased from three to five dimensions. Another point of interest here is that two Ss in the 4/8 proportion condition reported that they detected a difference in frequencies. Both of these thought there were more positive than negative instances.

Mean errors on pre-attainment blocks were plotted to compare performance, on both positive and negative instances, between Ss who claimed that they responded on the basis of differential frequencies and those who did not. These results are presented in Figure 9 for the 1/8, 2/8, and 3/8 conditions, combining the three- and five-relevant-dimension levels. At 1/8, the "frequency recognizers" were clearly superior on negative instances and showed greater superiority with trials. On positive instances their performance was worse than that of the "nonrecognizers" on the first half of the blocks, but the two groups were approximately the same on the last three blocks. At 2/8 the "frequency recognizers" were generally inferior to the "nonrecognizers" on negative instances, although this difference essentially disappeared in the last two blocks. The relationship between the two curves on positive instances is very similar to the relationship found at 1/8. At 3/8, the performance of the "frequency recognizers" is superior to the "nonrecognizers" on negative instances, but worse on positive instances. It should be noted that there were 17 "frequency recognizers" at 1/8, while there were only seven at 2/8, and three at 3/8.

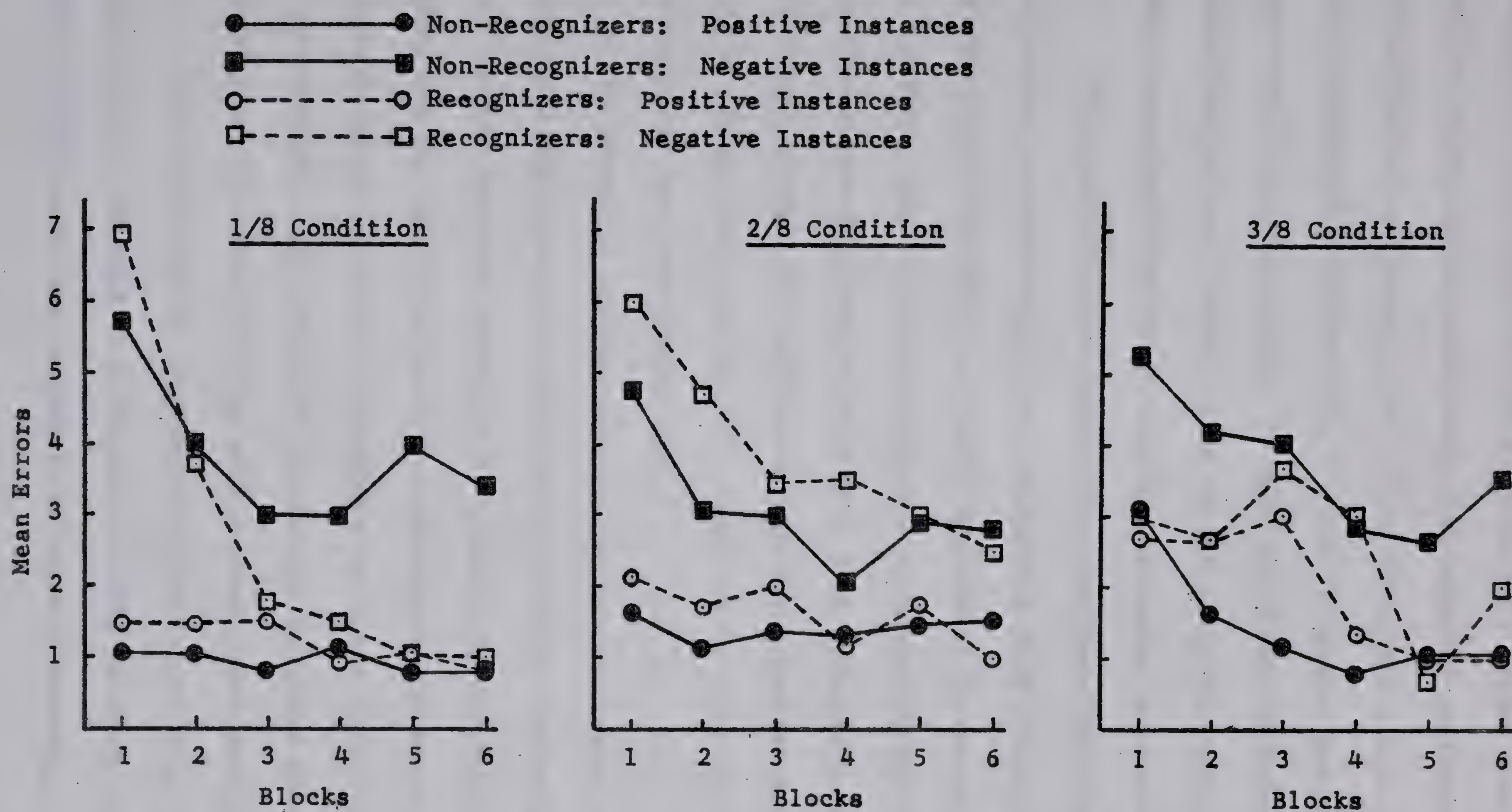

Fig. 9. Mean pre-attainment errors made by "frequency recognizers" and "nonrecognizers" on positive and negative instances across blocks in the 1/8, 2/8, and 3/8 proportion conditions.

These comparisons do not particularly clarify the picture, except at 1/8. There the two S groups were almost equal in numbers and the functions are consistent with expected performance based on probabilistic responding. At 2/8 the results are not consistent with probabilistic responding, but the "frequency recognizer" group consisted of only seven Ss. At 3/8, the three "frequency recognizers" theoretically had little to gain by frequency responding, but their performance was superior on negative instances and worse on positive.

To investigate the possibility that some individual Ss used maximizing as a strategy (i.e., responding 100% of the time with the category most frequently correct), individual S records were inspected. It was found that at 1/8, four Ss had response records indicating possible maximizing, but in each case this occurred on only one pre-attainment block. One additional S possibly maximized for one block at 2/8. There were none at 3/8. Thus, it is clear that a maximizing strategy was not used extensively.

To determine whether evidence of pre-attainment probability matching would show up in individual S records, the ratio of errors per block to expected errors under probability matching were plotted separately for each S in the 1/8 condition at the three- and five-relevant-dimension levels. Inspection of these curves yields no clearly interpretable indication of probability matching since most records showed great variability along with a general downward trend.

It is possible to view concept learning as an all-or-none phenomenon, as do Bower and Trabasso (1964), in which case pre-attainment performance is seen as devoid of concept learning; or else it may

be viewed as an incremental process, as assumed by Bourne and Restle (1959), in which pre-attainment responses are partly a function of concept learning. This distinction becomes important in the present study since pre-attainment responding based on differential frequencies is of particular interest, and therefore, assumptions about the composition of pre-attainment responses determine what is acceptable evidence of probabilistic responding.

Under the assumption of an all-or-none model, pre-attainment responding based on probability learning should show up in differences in button pushing between proportion curves. Under the assumption of an incremental process, on the other hand, it must be assumed that concept learning is intertwined with any probabilistic responding that may be present pre-attainment. In this case, the effects of probability learning are apparent only through inference rather than by direct observation.

To compare group mean responses with what might be expected if Ss were responding on the basis of probability matching, mean "plus" button responses were plotted as a function of blocks prior to attainment of the concept. (It will be recalled that attainment is defined as one perfect block with not more than one error per block thereafter.) Figure 10 shows these results. Assuming an all-or-none model, pre-attainment responding should be at approximately chance level or a mean of eight "plus" button responses per block despite the proportion of positive instances if probability learning were not occurring. The differential deviation from chance obtained in the order of the four curves is perfectly correlated with the order expected if prob-

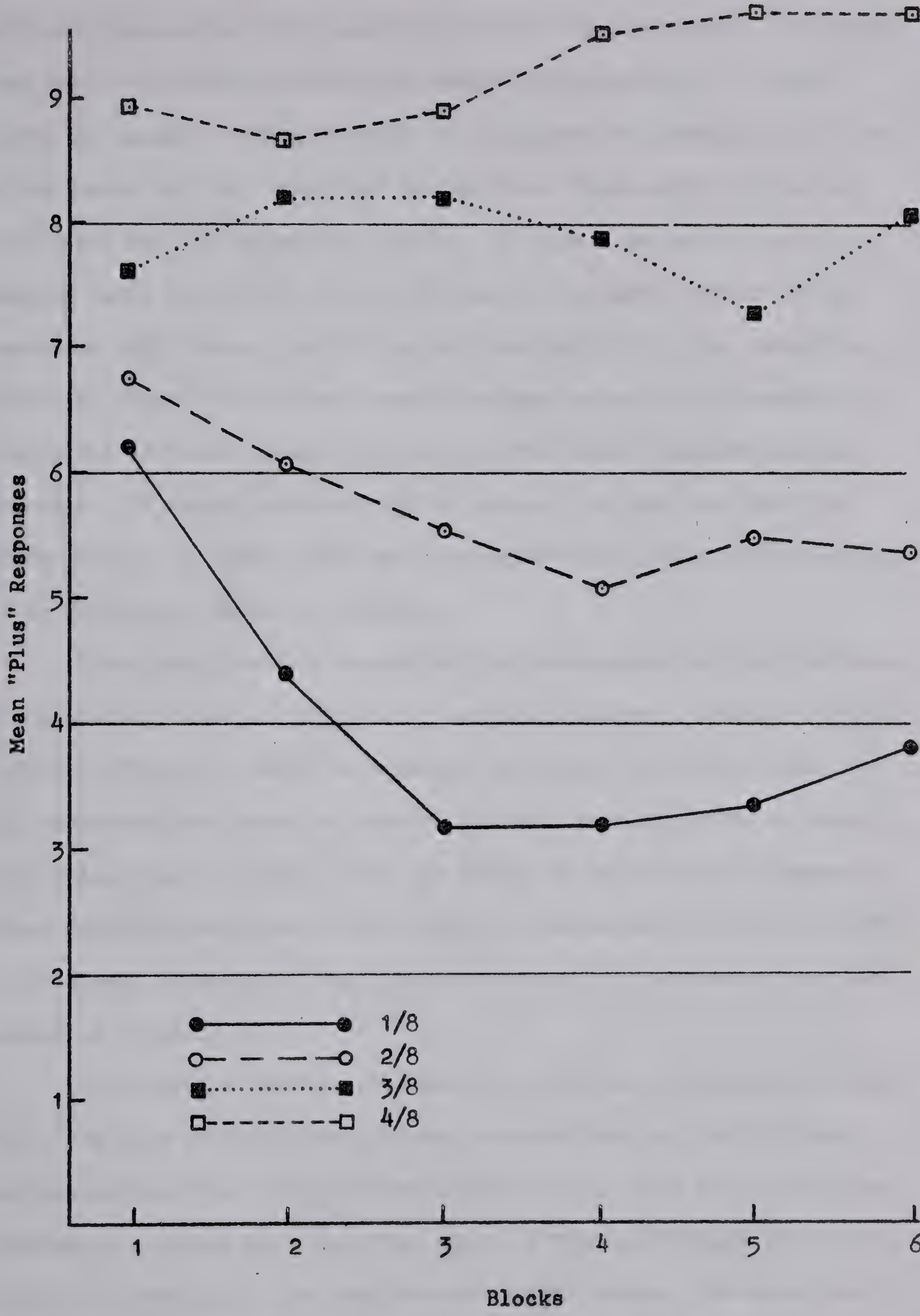

Fig. 10. Mean "plus" button responses made prior to attainment for each proportion condition plotted over blocks. The solid horizontal lines at 2, 4, 6, and 8 on the ordinate indicate the expected frequencies under probability matching for the 1/8, 2/8, 3/8, and 4/8 conditions, respectively.

abilistic responding were occurring. Generally, the curves are higher than would be expected under pure probability matching. In other words, Ss tended to respond prior to attainment by pressing the "plus" button more, and the "unmarked" button less, than would be expected. This holds for all proportion levels. It should be pointed out that because these data refer to pre-attainment responses, number of Ss decreased with blocks, and did so differentially with the proportion variable. Thus, the faster concept learners were not represented in the curves as blocks increased, leaving the slower learners and non-learners. Of major importance here, however, is the fact that the curves are in the order that would be expected in probability learning if an all-or-none model is assumed.

These data probably should not be interpreted as clear evidence of responding based on probability learning, however, because the four curves in Figure 10 would be expected to fall in this same order if Ss' responses were based on partial concept learning, with no probability learning involved. That is, either probabilistic responses or event matching responses should result in the order of the four curves as obtained, if concept identification occurs in increments over the course of trials.

To determine whether stationarity would be a reasonable assumption, the last pre-attainment blocks were omitted at the 4/8 level, and mean errors for this group were calculated. This was done to see whether or not the deviation from chance performance might be attributable to learning in the last pre-attainment block. The means over five blocks dropped from 5.4 errors on the first block, to 3.8 errors

on the second, and were at approximately 2.5 errors on the last three blocks. These findings do not support an all-or-none assumption in the present data, and deviations from chance responding should not therefore be interpreted as evidence of probability responding.

DISCUSSION

The major findings of the study may be briefly summarized as follows:

1. Error Analyses

a. Total errors increased as a function of proportion of positive instances from 1/8 to 3/8, but were lowest at 4/8.

b. Total errors increased as a negatively accelerated function of size.

c. Size and proportion were found not to interact.

d. Errors as a function of blocks decreased as a smooth logarithmic function.

2. Concept Attainment

a. The number of Ss attaining the concept increased as a function of proportion of positive instances in the order 3/8, 2/8, 1/8, 4/8.

b. The number of attainers decreased as size increased.

3. Positive and Negative Instances

a. Percent positive instance errors decreased as proportion increased. Negative instance errors were relatively constant at all proportions except 3/8, where errors were higher. At 1/8 and 2/8, percent positive errors exceeded negative, and at 3/8 and 4/8 the relationship was reversed.

b. When an equal number of positive and negative instances were presented, errors were greater on negative instances. The difference between positive and negative errors increased as size in-

creased.

c. The influence of negative instances on positive instance performance was not found to be facilitative, and perhaps served to interfere with optimal performance. The influence of positive instances on negative instance performance was somewhat uneven, but may be regarded as having been facilitative if probability learning is assumed to have been present in some conditions.

4. Pre-attainment Responding

a. Mean errors over blocks prior to concept attainment differed for the proportion groups in similar fashion to the overall mean error functions.

b. The number of Ss reporting that they responded on the basis of differential frequencies was greatest at the three-relevant-dimension level, with none at the one-relevant-dimension level. The number of such "recognizers" varied as a function of the frequency differential.

c. "Frequency recognizers" were superior to other Ss on negative instances pre-attainment in the 1/8 and 3/8 groups, but not in the 2/8 group. In all three groups "frequency recognizers" made more mean positive errors than did other Ss.

d. Probability maximizing was evident as a strategy in few cases, if at all. There was also little direct evidence of probability matching.

e. Pre-attainment "plus" button responding was proportional to the differential category frequencies.

The fact that the variable of proportion of positive instances was statistically significant confirms the prenotion that this is a variable of importance in concept identification performance. This demonstration, together with the fact that studies often hold this variable constant, leads to the conclusion that performance levels obtained in such studies may well be at least partly a function of the constant proportion level selected, and the results should be interpreted accordingly. Schvaneveldt (1965) has recently demonstrated that this variable is important in a trials to criterion task as well.

The present study found that performance deteriorated as proportion of positive instances increased, with the exception of the condition in which the frequencies of positive and negative instances were equal. Except for this condition, the present results are in direct contrast to Schvaneveldt's findings which showed that performance improved as proportion of positive instances was increased. This difference is confirmation of the writer's prediction of the results that should occur in a fixed trials situation as opposed to a trials to criterion task. This prediction was based on the reasoning that in a fixed trials situation Ss may be able to utilize differential frequencies with some effect upon performance measured in terms of total errors, whereas they would not be able to do so to the same extent in a trials to criterion situation.

The most likely explanation for the overall error results on the proportion variable lies in the presence of two oppositional factors. On the one hand, it might be expected that concept learning would be better as the proportion of positive instances is

increased. On the other hand, as the frequency disparity between the categories of positive and negative instances is increased in the direction of fewer positive instances, performance should be enhanced by the opportunity for probability learning and responding on this basis.

The various analyses of error data are generally consistent with this point of view. This interpretation would suggest that in the overall error results the 1/8 and 2/8 groups are lowered in errors as a result of probabilistic responding. It is not unreasonable to assume that probabilistic responding had little effect on 3/8 error scores since Ss in this condition had relatively little to gain over chance performance by responding on the basis of frequencies as compared with the 1/8 and 2/8 Ss.

The curves showing percent errors on positive and negative instances by the different proportion conditions are also consistent with this explanation. These results indicate that percent negative instance errors are relatively low at 1/8 and 2/8, while positive instance errors at these same levels are higher than at 3/8 and 4/8. This is what would be expected under probabilistic responding. That is, Ss should show lower percent errors on negative instances in proportion to the magnitude of the disparity, but should show an elevation in percent positive instance errors. If one assumes that more event matching is occurring on positive instances, however, then this effect of probabilistic responding may not be as evident in positive instance error results.

The data on the first 12 positive instances and on the first 48 negative instances are also consistent with the probability

learning explanation. When the data on the first 12 positive instances are considered, it appears that frequency responding likely exerted an effect on these data in that errors generally increased as the number of interspersed negative instances increased. This is consistent with the expectation that responding to the category with the greatest frequency would tend to elevate positive instance errors as the frequency disparity is increased. The data on the first 48 negative instances also support this view in that the 1/8 and 2/8 groups fell midway between the 3/8 and 4/8 groups in errors, suggesting that the influence of number of interspersed positive instances was obscured by responding on the basis of the learned differential frequencies.

The probabilistic responding explanation of the data is consistent with the findings of Mandler, Cowan, and Gold (1963). They found that Ss probability matched prior to concept attainment. Their finding that this occurred to a lesser extent as the concept problem became more difficult is partially consistent with the present results in that verbal reports of Ss suggest a greater amount of frequency responding at the three-relevant-dimension level than at the five-relevant-dimension level, but none at the one-relevant-dimension level.

Under pure probability matching in the present study, Ss in the 1/8 condition would be expected to make 3.5 errors per block; in the 2/8 condition, 6.0 errors per block; and in the 3/8 condition, 7.5 errors per block. Thus, it is clear that the 3/8 condition Ss had almost nothing to gain by probability matching. The 2/8 condition Ss

had a bit more to gain, and the 1/8 condition Ss had a great deal to gain over random responding. Under probability maximizing, the order is the same, but all three of these groups would have been able to benefit even more, as compared with random responding. If the results are viewed in light of expected benefit they are consistent.

Attempts at directly observing probabilistic responding in the data were largely unsuccessful probably because of the complexity of conceptual functioning. While some Ss claimed to have responded on the basis of differential frequencies, comparison of their pre-attainment error means with the other Ss at that proportion condition did not provide a uniformly clear substantiation of verbal reports, except at 1/8 and 3/8. The "frequency recognizers" at 2/8 did not perform as would be expected. Inspection of individual error records for probability maximizing and probability matching yielded little clear evidence of either type in its pure form. Pre-attainment errors and pre-attainment "plus" button responses both showed results consistent with probabilistic responding, but did not provide direct evidence since an assumption of "all-or-none" learning cannot be made. Thus, in the present data, deviation of pre-attainment responding from chance is not adequate as direct evidence of probability learning.

The absence of direct evidence does not rule out the likelihood that the major results are best accounted for on the basis of differential frequency responding. It is highly likely that the size of the blocks, i.e., 16 instances, constitutes an insensitive indicator of probability matching or maximizing, particularly since a solvable problem always existed. Subjects could very easily respond on the

basis of probabilities for a portion or portions of individual blocks, but respond on the basis of hypotheses in other portions of the blocks. A S may well have adopted the strategy, "if in doubt, press the unmarked button". Indeed, this is probably the optimal approach since it would constitute a combination of partial concept learning with probability maximizing. If this were in fact the case, neither probability maximizing nor probability matching would be directly observable in the data.

The results of concept attainment vs. nonattainment as related to the different proportions of positive instances is perfectly consistent with the overall error results in that the greatest number of attainers occurred in the 4/8 condition, followed by 1/8, 2/8, and 3/8 in that order. At first glance this congruence would seem to be as expected. However, the explanation tendered for the error results does not necessarily apply in the case of concept attainment. Although probabilistic responding could account for differential error rates between the proportion conditions, it is not obvious that this would also lead to differential frequencies of attainment. On the contrary, it might be expected that attainment would be more consistent with Schvaneveldt's (1965) results such that the frequency of attainment would increase as proportion of positive instances is increased.

It was considered that the reason for this consistency between measures might be that when Ss attained the concept their errors dropped to essentially zero. Thus, differential rates of concept attainment could account for differences in mean errors without there being any differences in mean errors prior to attainment. Graphs of

mean pre-attainment errors over blocks were shown to vary as a function of proportion of positive instances in similar fashion to the overall mean errors presented among the results of the overall error analysis. This suggests that the overall results did not arise from differential rates of attainment.

Two other alternatives suggest themselves as possibilities in an effort to explain the congruence between error and attainment results. The first is that each constitutes an independent assessment of performance and that they both result from the same phenomenon. Since attainment should not be directly affected by presolution probabilistic responding, but instead should reflect learning or identification of the concept, then its correlation with error results suggests that perhaps the error results too reflect mainly concept learning and not probabilistic responding as has been suggested. If this is the case then the results of the present study are inconsistent with other results involving proportions of positive and negative instances. The research by Hovland and Weiss (1953) and by Schvaneveldt (1965) would lead to the prediction that learning, and hence performance, would improve as the proportion of positive instances is increased. The superiority of the 4/8 condition is the only finding which is consistent with this prediction, while the other results are directly opposite. Thus, the difference between the present results and earlier findings argues strongly against this proposition.

This leads to the remaining alternative, namely, that perhaps differences in error performance in some way led to differential frequencies of attainment. Eventual attainment of the concept may be

facilitated by greater accuracy throughout the course of the problem regardless of the basis for such accuracy. Consistent with a conditioning approach to concept learning, it is possible that the greater the number of positively reinforced responses to a given type of instance, the stronger will be the association of that type of instance and its correct response. Despite the factors which preceded the making of correct responses to particular types of instances, viz. probability learning, reinforcement of those responses in the form of informative feedback should lead to differential conditioning of responses to types of stimuli. Thus, this explanation suggests that as the number of correct responses increased, so too did the number of Ss who learned the differential response to the different types of stimuli to a criterion of 100% correct responding. If this explanation is valid, then the two primary measures--errors and attainment--are not independent. This is also consistent with the usual finding of high intercorrelations between measures (Bourne, 1966).

An additional finding of interest is that errors increased as both a linear and quadratic function of size. This negatively accelerated function confirms the earlier finding by Wargo (1960). Both studies used similar methodology and ran Ss for a fixed number of instances. By the same token, the present results are inconsistent with those of Bulgarella and Archer (1962), and with those of Walker and Bourne (1961). Here again there is the suggestion that when running Ss for a fixed number of trials, the differential frequencies serve to facilitate error reduction to a greater extent as size is

increased, producing the negatively accelerated function.

If the absence of a significant size x proportion interaction is considered, however, then the validity of this suggestion is questionable. The proportion functions across size must be regarded as not significantly deviating from parallel. Described another way, the rate of increase in errors as size was increased did not significantly differ for the proportion conditions. Since the proportion conditions permit different degrees of probabilistic responding, deviation from parallel would be expected if the major cause of the negatively accelerated function was attributable to differential degrees of probabilistic responding as size was increased. These data, then, do not support the proposition that the differential frequencies will serve to reduce errors to a greater extent as size is increased. It is of interest to note, however, that in the 4/8 condition, where probabilistic responding was impossible, errors appear to have increased in a linear fashion as a function of size, while the other three functions were clearly curvilinear. This lends some support to the proposition of differential rates of error increase. Overall, however, the data relating to this point do not strongly support the original proposition.

Results of the analysis at the 4/8 proportion level on types of instances show that performance on positive instances at this level was superior to negative instance performance, and that this differential increased with size. This may be due to the fact that there were more different kinds of negative instances than of positive instances, except at the one-relevant-dimension level. The number of

kinds of negative instances increased with size, while there were always only two kinds of positive instances.

Another finding worthy of mention is that the pre-attainment "plus" button response curves show evidence of a positive response bias. That is, Ss tended to make "plus" responses more frequently than would be expected on the basis of any factors identified as operating thus far. This may be a function of the instructions, or perhaps a function of the manner of designating the categories, i.e., "plus" vs. "unmarked".

In conclusion, the major findings of the present study have serious implications for concept learning research. The finding that different category frequencies led to differences in performance suggests that when a fixed number of trials is employed, the proportion of positive and negative instances of the concept is a factor to be considered in that results may vary as a function of any disparity, very likely due to probabilistic responding. Some studies which have employed such methodology might be re-examined on the basis of the present demonstration.

The present results suggest that the use of errors as the measure of performance in a fixed trials concept identification task may yield results based on factors inherent in this methodology unless steps are taken to guard against these special effects. Indeed, other experimental situations outside concept research may exist where probabilistic responding is possible. The researcher should be aware of the presence of this possibility. So, too, should the educator be cognizant of this in connection with the communication of concepts and

the assessment of conceptual learning.

In general, it would seem advisable that the researcher or educator employ equal frequencies of positive and negative instances so as to eliminate the possibility that probabilistic responding will affect performance when such responding is not of interest.

If the probability learning explanation for the results of the present study are valid, then they lead to interesting implications for an understanding of human conceptual functioning. They suggest that human beings may well utilize differential frequencies when a category response is demanded in a situation. It would appear that probabilistic responding might constitute a well learned approach to problem situations among adults to be employed where problem-relevant information has not yet been learned. But, perhaps most importantly, if this is the seemingly irrelevant basis upon which correct pre-solution responses are sometimes made, the eventual attainment of the concept may not greatly suffer since concept attainment was greatest in those conditions where pre-attainment correct responses were greatest. Since some of these conditions were held to have low error rates because of probability responding, then it is suggested that high rates of correct responding may lead to concept solution.

SUMMARY

The present study investigated the effects of proportion of positive instances (1/8, 2/8, 3/8, and 4/8), and concept size (one, three, and five nonredundant relevant dimensions) on errors in a two choice, visual, conjunctive concept identification task where Ss were presented a fixed number of 96 instances.

Mean total errors were found to increase as proportion of positive instances increased in the range 1/8, 2/8, and 3/8, but were lowest at 4/8. Mean total errors increased as a negatively accelerated function of size. The interaction between these two variables was not statistically significant. Errors per block of 16 instances were found to decrease in a smooth logarithmic manner as blocks increased. *Post hoc* criterion analysis of concept attainment showed results consistent with overall error scores. Analysis of positive vs. negative instance errors at the 4/8 proportion condition showed that performance on positive instances was significantly superior, and this difference increased as size increased.

The results of the study suggest that Ss in the 1/8 and 2/8 proportion conditions reduced pre-attainment errors by responding on the basis of learned frequencies. Frequency of concept attainment was explained as resulting from the number of correct reinforced responses made prior to attainment.

References

Archer, E. J. Concept identification as a function of obviousness of relevant and irrelevant information. _J_. _exp_. _Psychol_., 1962, 63, 616-620.

Archer, E. J., Bourne, L. E. Jr., & Brown, F. G. Concept identification as a function of irrelevant information and instructions. _J_. _exp_. _Psychol_., 1955, 49, 153-164.

Bourne, L. E. Jr. Effects of delay of information feedback and task complexity on the identification of concepts. _J_. _exp_. _Psychol_., 1957, 54, 201-207.

Bourne, L. E. Jr. _Human conceptual behavior_. Boston: Allyn and Bacon, 1966.

Bourne, L. E. Jr., & Pendleton, R. B. Concept identification as a function of completeness and probability of information feedback. _J_. _exp_. _Psychol_., 1958, 56, 413-420.

Bourne, L. E. Jr., & Restle, F. Mathematical theory of concept identification. _Psychol_. _Rev_., 1959, 66, 278-296.

Bower, G., & Trabasso, T. Concept identification. In R. C. Atkinson (Ed.), _Studies in mathematical psychology_. Stanford: Stanford University Press, 1964. Pp. 32-96.

Brown, F. G., and Archer, E. J. Concept identification as a function of task complexity and distribution of practice. _J_. _exp_. _Psychol_., 1956, 52, 316-321.

Bruner, J. S., Goodnow, Jacqueline J., & Austin, G. A. _A study of thinking_. New York: Wiley, 1956.

Bulgarella, R. G., & Archer, E. J. Concept identification of auditory stimuli as a function of amount of relevant and irrelevant information. _J_. _exp_. _Psychol_., 1962, 63, 254-257.

Freibergs, Vaira, & Tulving, E. The effect of practice on utilization of information from positive and negative instances in concept identification. _Canad_. _J_. _Psychol_., 1961, 15, 101-106.

Gaito, J., & Wiley, D. E. Univariate analysis of variance procedures in the measurement of change. In Chester W. Harris (Ed.), _Problems in measuring change_. Madison: University of Wisconsin Press, 1963. Pp. 60-84.

Grant, D. A. Analysis of variance tests in the analysis and comparison of curves. _Psychol_. _Bull_., 1956, 53, 141-154.

Heidbreder, Edna. The attainment of concepts: I. Terminology and methodology. J. gen. Psychol., 1946, 35, 173-189.

Heidbreder, Edna. The attainment of concepts: III. The process. J. gen. Psychol., 1947, 24, 93-138.

Hovland, C. I., & Weiss, W. Transmission of information concerning concepts through positive and negative instances. J. exp. Psychol., 1953, 45, 175-182.

Hull, C. L. Quantitative aspects of the evolution of concepts. Psychol. Monogr., 1920, No. 123.

Hunt, E. B. Concept learning: An information processing problem. New York: Wiley, 1962.

Kendler, Tracy S. Concept formation. In Annu. Rev. Psychol., 1961, 12, 447-472.

Kendler, H. H. The concept of the concept. In Arthur W. Melton (Ed.), Categories of human learning. New York: Academic Press, 1964. Pp. 211-236.

Lindquist, E. F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953.

Mandler, G., Cowan, P. A., & Gold, C. Concept learning and probability matching. J. exp. Psychol., 1964, 67, 514-522.

Miller, G. A. What is information measurement? Amer. Psychologist, 1953, 8, 3-11.

Schvaneveldt, R. W. Concept identification as a function of probability of positive instances and amount of relevant information. Paper read at Midwestern Psychological Association Meeting, 1965.

Smoke, K. L. An objective study of concept formation. Psychol. Monogr., 1932, 42, No. 4.

Smoke, K. L. Negative instances in concept learning. J. exp. Psychol., 1933, 16, 583-588.

Underwood, B. J. An orientation for research on thinking. Psychol. Rev., 1952, 59, 209-220.

Walker, C. M., & Bourne, L. E. Jr. Concept identification as a function of amounts of relevant and irrelevant information. Amer. J. Psychol., 1961, 74, 410-417.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.

Wargo, D. G. Concept identification as a function of relevant and irrelevant information. Unpublished master's thesis, University of Wisconsin, 1960.

APPENDIX I

Instructions to Subjects

"In this experiment you are going to see a series of 96 slides projected on the screen in front of you. The slides will differ from each other in certain ways. Your task will be to assign each of these slides to one of two categories represented by the two buttons on the panel in front of you. It will be up to you to figure out what it is about some of the slides that makes them go into the category marked by the plus sign. After each slide is presented to you, you are to decide whether or not it belongs in the plus category. If it does, then press the button marked by the plus sign. If you don't think it belongs in that category, then press the unmarked button.

"As soon as you press one of the buttons the light above the button which was the correct button will go on. So, if the light goes on over the button you pressed, then you made a correct response. If the light goes on over the other button, then you made an error. The light is simply to tell you which button you should have pressed. Try to make as many correct responses as you can.

"A slide will stay on the screen until both of you have pressed a button. Then the next slide will come on. Press the button you decide upon as soon as possible, but being correct is more important than speed. Once you press a button you cannot change your mind, so be sure before you respond.

"You are not competing with each other. It is merely more efficient to run more than one person at a time. Each of your decisions are being recorded separately.

"Now remember, it is up to you to figure out what it is about some of the slides that puts them in the plus category. For each slide, decide what category it goes into and press the corresponding button. The light will tell you which button was the correct one. Your score will be the total number of correct responses that you make."

APPENDIX II

SUBJECT INFORMATION QUESTIONNAIRE

Name______________________________________Sex________Age_________

Faculty___________________________________Year____________

1. Were you able to figure out what it was that made some of the slides go into the "plus" category?_______________

Describe the slides that went into the "plus" category:_____________

____

____

2. After you figured it out, did you make any more errors?_________

If you did, explain why:____

____

3. Before you figured it out did you press one of the buttons more often than the other?________________

If you did, explain why:____

____

____

4. Additional comments:____

____

____

APPENDIX III

Raw Data: Errors Per Block For Each S

One Relevant Dimension

Problem	Proportion	S	Male Blocks 1	2	3	4	5	6	Female Blocks 1	2	3	4	5	6
A	1/8	1	7	5	0	0	0	0	2	0	0	0	0	0
		2	4	0	0	0	0	0	7	5	2	0	0	0
		3	11	10	8	7	6	9	1	0	0	0	0	0
		4	2	2	1	1	0	0	2	0	0	0	0	0
	2/8	1	1	0	0	0	0	0	0	0	0	0	0	1
		2	4	1	0	0	0	0	2	0	0	0	1	0
		3	2	0	0	0	0	0	2	0	0	0	0	0
		4	3	0	1	1	0	0	2	0	0	0	0	0
	3/8	1	2	0	0	0	0	0	2	0	0	0	0	0
		2	6	5	3	4	0	0	1	0	0	0	0	0
		3	2	0	0	0	0	0	3	0	0	0	0	0
		4	1	1	0	0	0	0	7	5	7	7	0	0
	4/8	1	0	0	0	0	0	0	2	0	0	0	0	0
		2	1	0	2	1	1	0	3	1	0	3	0	1
		3	2	0	0	0	0	0	3	0	0	0	0	0
		4	3	0	0	0	0	0	8	5	5	1	4	0
B	1/8	1	2	0	0	0	0	0	1	0	1	0	0	0
		2	2	0	0	0	0	0	4	1	0	0	0	0
		3	2	0	0	0	0	0	2	0	0	0	0	0
		4	3	0	1	0	0	2	2	0	0	0	0	0
	2/8	1	9	2	0	0	0	1	2	0	0	0	0	0
		2	4	0	0	0	0	0	8	4	1	0	0	0
		3	1	0	1	0	0	0	2	0	0	0	0	0
		4	4	1	2	1	0	1	3	0	0	0	0	0
	3/8	1	4	1	0	0	0	1	2	0	0	0	0	0
		2	2	0	0	0	0	0	6	11	9	11	5	3
		3	5	0	0	0	0	0	5	0	0	2	0	0
		4	9	8	8	0	0	1	11	4	1	1	0	0
	4/8	1	3	0	0	0	0	0	3	1	0	0	0	0
		2	1	0	0	0	0	0	4	0	0	0	0	0
		3	3	1	0	0	1	0	1	0	0	0	0	0
		4	2	0	0	0	0	0	4	0	0	0	0	0

Three Relevant Dimensions

			Male						Female					
	Propor-		Blocks						Blocks					
Problem	tion	S	1	2	3	4	5	6	1	2	3	4	5	6
C	1/8	1	5	8	6	2	6	4	5	3	2	0	0	0
		2	10	5	1	2	0	1	8	6	3	0	0	0
		3	7	3	2	1	1	0	9	6	4	3	1	0
		4	7	4	2	2	3	1	12	7	5	2	4	4
	2/8	1	9	6	9	5	3	4	6	1	1	2	0	0
		2	6	5	2	3	6	5	7	4	0	0	0	0
		3	11	6	7	4	7	3	6	1	3	4	0	0
		4	6	1	1	0	0	0	6	3	9	5	9	4
	3/8	1	3	3	0	0	0	0	10	5	8	5	6	4
		2	9	4	8	4	4	7	6	7	6	3	0	3
		3	4	0	0	0	0	0	6	2	1	1	1	2
		4	8	5	10	7	9	9	6	1	2	1	1	1
	4/8	1	7	8	1	2	0	0	8	4	4	1	2	1
		2	6	4	0	2	0	0	6	4	4	2	3	4
		3	7	1	1	0	0	0	3	1	0	0	0	0
		4	6	2	1	2	1	0	4	3	1	3	1	0
D	1/8	1	11	6	0	0	0	0	8	4	1	0	1	0
		2	10	7	0	0	0	0	7	3	0	1	0	0
		3	7	4	3	1	4	1	8	4	3	2	1	0
		4	10	5	7	3	2	0	8	3	0	0	0	0
	2/8	1	5	4	4	4	0	0	8	6	5	4	0	0
		2	9	9	12	5	7	9	4	7	3	2	0	0
		3	10	9	6	5	5	7	7	3	3	1	0	0
		4	3	2	0	1	0	0	6	0	0	0	0	0
	3/8	1	12	7	7	7	3	4	5	9	7	3	6	3
		2	11	10	7	8	2	4	11	5	7	3	0	5
		3	8	8	4	4	6	5	13	6	5	2	1	2
		4	4	2	3	1	0	0	10	2	4	0	0	1
	4/8	1	4	4	2	0	0	0	5	4	2	1	0	0
		2	5	0	0	2	0	0	4	0	1	3	0	0
		3	4	2	0	0	0	1	2	0	0	0	1	1
		4	4	1	0	0	0	0	2	0	0	0	0	1

Five Relevant Dimensions

			Male						Female					
	Propor-		Blocks						Blocks					
Problem	tion	S	1	2	3	4	5	6	1	2	3	4	5	6
E	1/8	1	9	11	6	4	6	5	2	1	0	1	0	0
		2	7	5	4	2	2	3	10	5	2	1	1	0
		3	5	5	1	1	1	0	6	3	0	0	0	0
		4	3	0	0	0	0	0	9	8	5	5	5	4
	2/8	1	6	5	3	3	1	0	7	2	4	4	2	0
		2	8	4	8	4	8	6	7	6	9	3	7	5
		3	8	9	4	5	6	1	8	5	1	0	0	0
		4	3	0	0	0	0	0	5	3	1	0	0	0
	3/8	1	7	6	5	2	0	0	6	11	6	4	6	6
		2	7	8	7	9	8	12	5	8	12	9	4	5
		3	8	11	7	3	2	0	4	5	6	2	2	3
		4	8	3	4	4	8	4	9	5	3	4	1	1
	4/8	1	4	0	0	0	1	0	2	2	0	0	0	0
		2	5	0	0	0	0	0	10	0	0	0	1	0
		3	4	3	4	1	0	0	7	1	4	2	3	2
		4	5	6	6	3	3	1	7	4	2	0	2	0
F	1/8	1	9	11	7	10	9	6	8	4	1	0	1	0
		2	7	6	4	3	2	1	7	10	7	6	2	2
		3	7	6	6	7	3	1	10	3	2	0	0	0
		4	8	1	1	0	0	0	6	2	3	1	0	0
	2/8	1	9	6	5	3	2	1	6	6	6	6	1	1
		2	7	10	10	8	6	3	6	1	8	6	5	9
		3	9	9	3	1	0	1	8	5	1	0	0	2
		4	6	3	1	0	0	0	5	1	1	2	1	2
	3/8	1	7	4	3	1	2	0	10	4	6	6	5	8
		2	9	3	6	2	2	6	13	9	7	3	5	3
		3	9	7	5	3	4	5	12	5	2	1	3	2
		4	11	7	1	0	0	0	8	7	2	2	0	2
	4/8	1	3	6	2	2	4	1	5	8	6	7	4	1
		2	5	3	2	3	1	2	9	5	2	7	4	2
		3	3	3	1	2	2	0	1	3	0	0	2	1
		4	6	6	2	1	1	1	13	4	2	0	2	2